

Introduction to Path Analysis

Hsueh-Sheng Wu
CFDR Workshop Series
November 18, 2019

BGSU



Center for Family and
Demographic Research

Outline

- What is path analysis?
- Structural Equation Modeling (SEM)
- Steps of using SEM in Stata to fit path models
- Choice between SEM and GSEM
- Estimation methods
- Model fit statistics
- Model modification
- Examples of using Stata to run path analysis
- Strengths and limitations of using SEM in fitting path models
- Conclusions

BGSU



Center for Family and
Demographic Research

What Is Path Analysis?

- Path Analysis is a statistical technique for examining relations among observed variables
- In 1918, Sewall Wright, a geneticist, was the first scholar using path diagrams to show possible causal relations between variables
- Regression techniques such as OLS regression or logistic regression examine the relations between multiple predictors and one outcome variable, which is a simple form of path model
- Regression analysis can be expanded to become complex path models by including additional dependent variables, mediating variables, and moderating variables
- Complex path models can only be estimated within the framework of Structural Equation Modeling

BGSU

Structural Equation Modeling (SEM)

- Structural Equation Modeling allows researchers to hypothesize a set of relations between variables and test them on observed data
 - Variables , paths, and effects in SEM
 - Types of SEM models
 - Stories that SEM tells
 - Nested vs not-nested models
 - Parameters in SEM

Variables, Paths, and Effects in SEM

- Variables in SEM
 - Measured variable
 - Latent variable
 - Exogenous variable
 - Endogenous variable
 - Residuals in endogenous variable
 - Disturbance in exogenous variable
- Paths linking two variables
 - A single-headed arrow: one variable predicts the other variable
 - A double-headed arrow: two variables are correlated with each other
 - No path means no direct relation between two variables
- Effects of one variable on another variable
 - Direct effect
 - Indirect effect
 - Total effect
- Nested vs Not-nested model
- Recursive vs. Nonrecursive model

Types of SEM Models

- Measurement model
 - Factor analysis
 - Latent class analysis
 - Latent profile analysis
 - Multi-group and multi-level SEM model
- Structural model
 - Various regression analysis
 - Regression with mediators
 - Multi-group and multi-level SEM model
- Full SEM model
 - Combination of measurement model and structural model

BGSU



Center for Family and Demographic Research

Stories That SEM Tells

- SEM basically tells at least one of the following three stories:
 - What is the relation between variables?
 - Is this relation mediated by another variable?
 - Is this relation modified by another variable?
- SEM tells these stories through a series of proposing, testing, and comparing models

Stories That SEM Tells (Cont.)

- Lin, I-Fen and Hsueh-Sheng Wu. 2014. “Activity Limitations, Use of Assistive Devices or Personal Help, and Well-Being: Variation by Education.” *The Journals of Gerontology, Series B: Psychological Sciences and Social Sciences* 69(7):16-25
- Theoretical relation between activity limitations and well-being among older adults
 - Activity limitation is negatively related to well-being of older adults. However, since well-being is a multi-dimension construct, activity limitation may have different relations with these constructs
 - With activity limitations, older adults may use assistive devices and/or personal help to cope with the limitations, which in turn influences their well-being
 - Older adults with different levels of education may differ in their abilities to adopt various coping strategies, as well as in the extent to which they can accept the stigma of losing independence that is symbolized by using these strategies
 - Therefore, the associations among activity limitations, use of assistive devices or personal help, and 3 dimensions of well-being (positive affect, self-realization, and self-efficacy) differ for older adults with different levels of education

Stories That SEM Tells (Cont.)

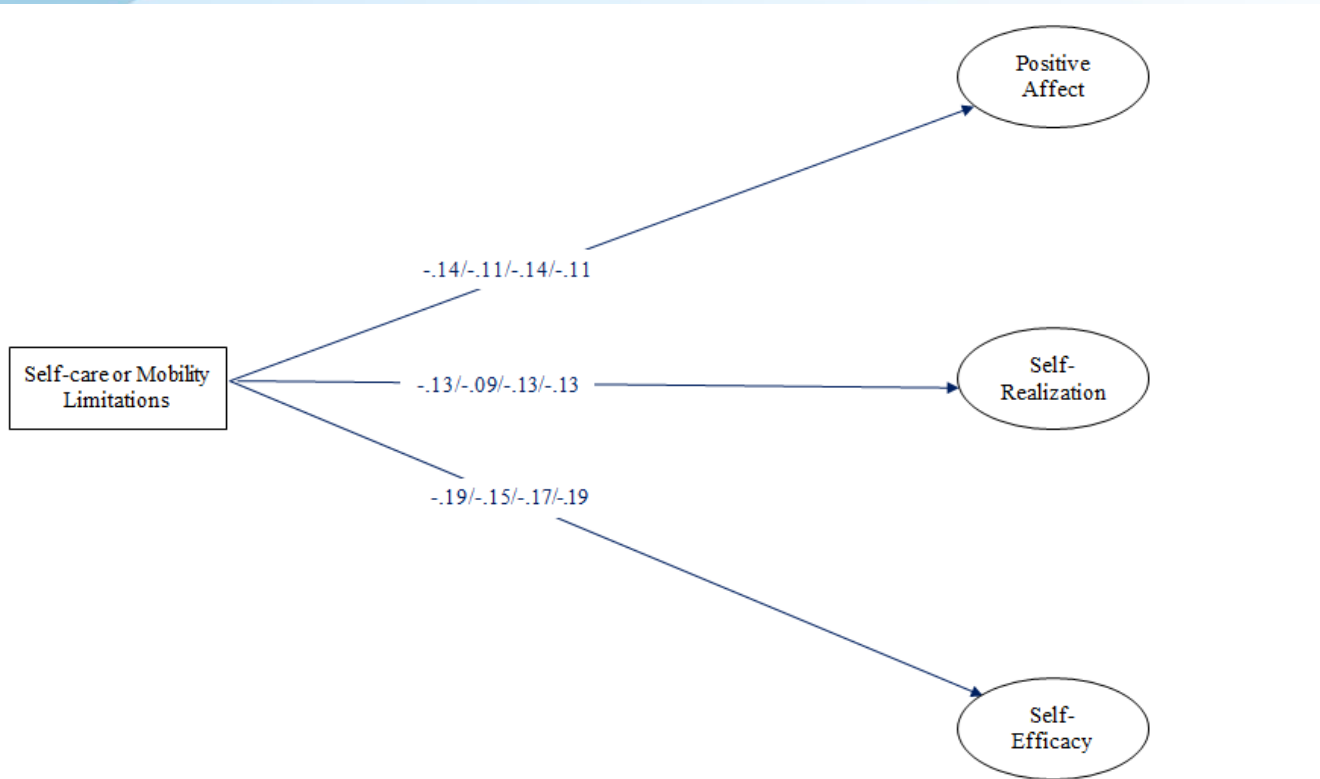


Figure 1. Unstandardized Path Coefficients for the Relations between Self-care or Mobility Limitations and Well-being among Older Adults

Note. To simplify the presentation, factor loadings and residual correlations are not shown. Numbers for each path indicate less than high-school education, high-school graduate, some college education, and a bachelor's degree or higher, respectively. All numbers are significantly different from zero at the $p < .05$ level.

Stories That SEM Tells (Cont.)

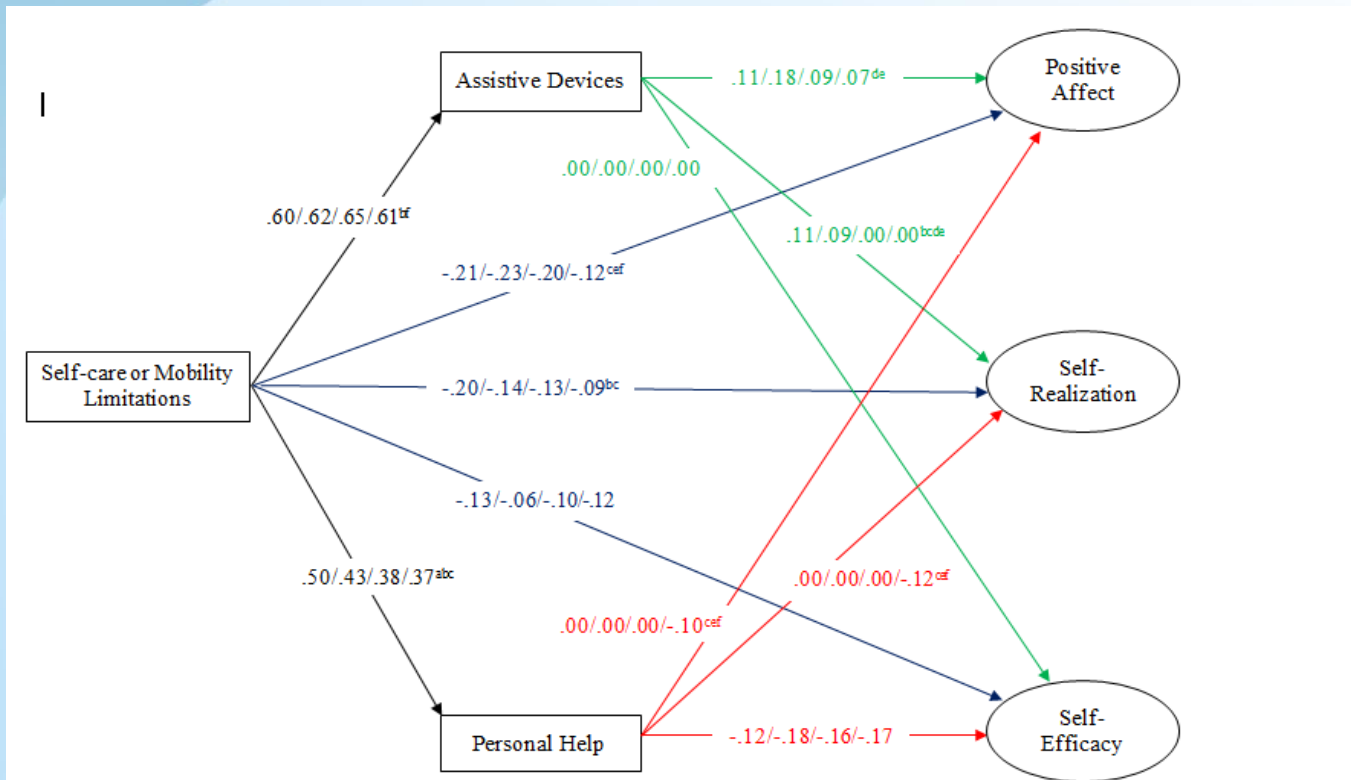
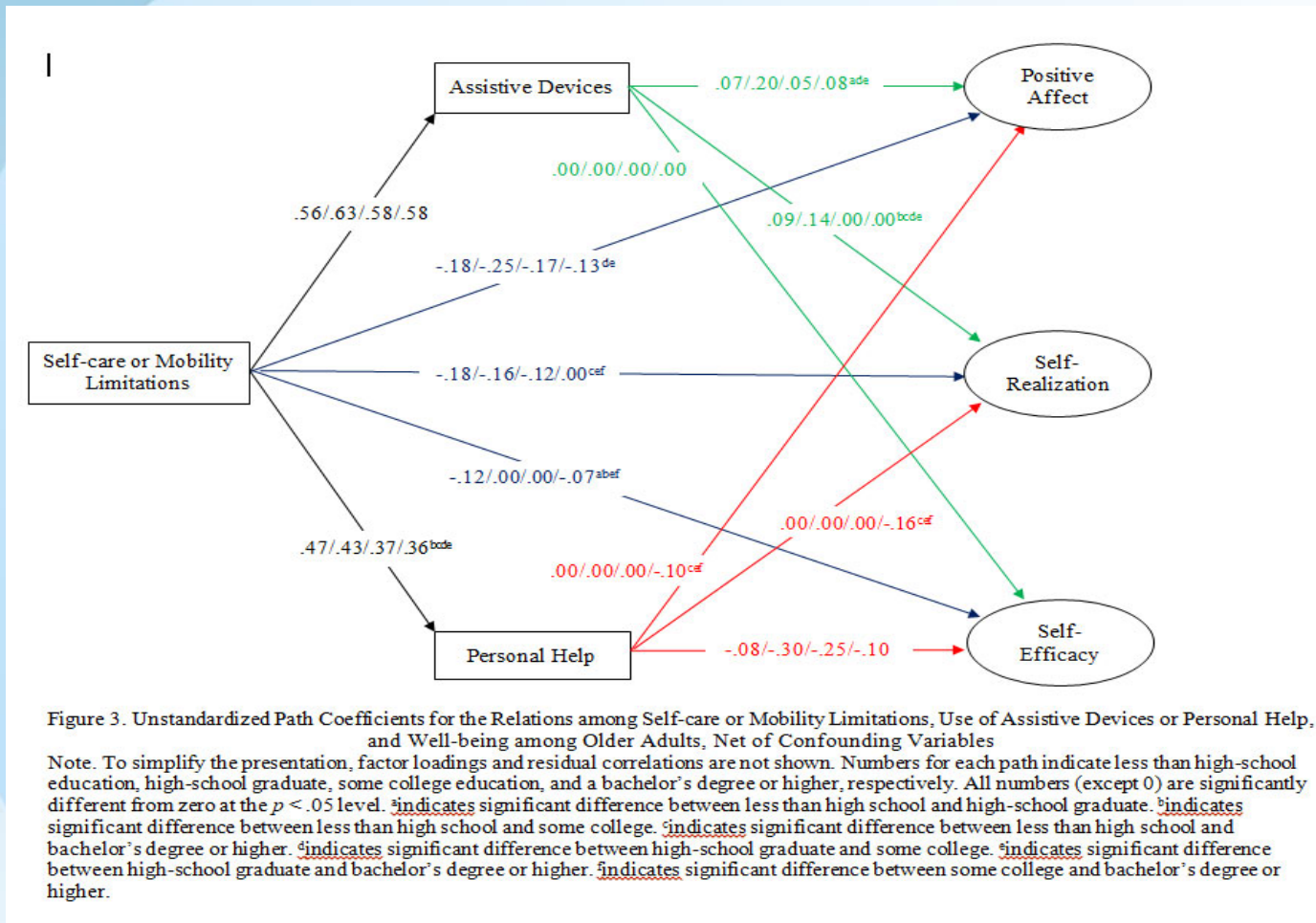


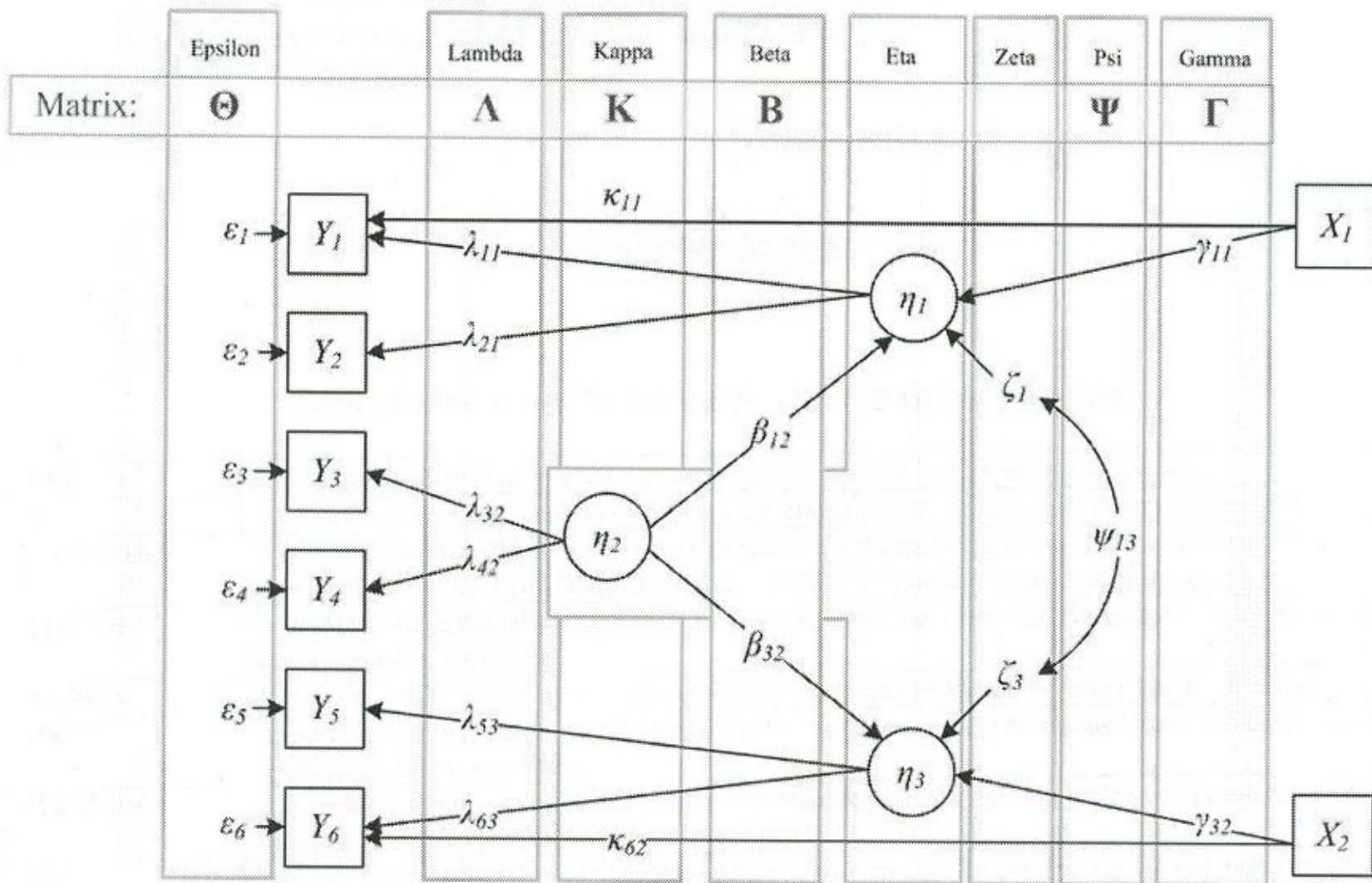
Figure 2. Unstandardized Path Coefficients for the Relations among Self-care or Mobility Limitations, Use of Assistive Devices or Personal Help, and Well-being among Older Adults

Note. To simplify the presentation, factor loadings and residual correlations are not shown. Numbers for each path indicate less than high-school education, high-school graduate, some college education, and a bachelor's degree or higher, respectively. All numbers (except 0) are significantly different from zero at the $p < .05$ level. ^aindicates significant difference between less than high school and high-school graduate. ^bindicates significant difference between less than high school and some college. ^cindicates significant difference between less than high school and bachelor's degree or higher. ^dindicates significant difference between high-school graduate and some college. ^eindicates significant difference between high-school graduate and bachelor's degree or higher. ^findicates significant difference between some college and bachelor's degree or higher.

Stories That SEM Tells (Cont.)



Parameters in SEM



Steps of Using SEM in Stata to Fit Path Models

- Theorize a SEM model
 - Develop a theoretically based model
 - Construct the SEM diagram
 - Convert the SEM diagram into a set of structural equations
- Fitting a series of SEM models
 - Check if the model is identifiable
 - Choose between SEM and GSEM commands
 - Determine the estimation method
 - Run the model and evaluate goodness of fit statistics
 - Modify the model by adding a path, dropping a path, fixing the coefficient of a path, or setting paths to be equal
 - Compare two models and decide if additional modification is needed
 - After the final model is decided, report the results

Steps of Using SEM in Stata to Fit Path Models (Cont.)

SEM Models are generally identifiable if the number of data points is smaller than the number of free parameters.

- In the measurement part:
 - When there is only one latent variable in the model, this latent variable needs to have three indicators and the error terms of these indicators are correlated
 - When there are more than one latent variable in the model, the three-indicator rule mentioned above still applies. The model is identifiable if a latent variable has only two indicators, the error terms of these two indicators are not correlated, and the latent variables are correlated.
- In the structural part:
 - When the structural part is recursive, meaning (1) none of the endogenous variables predicts each other and (2) no correlated errors between endogenous variables
 - When the structural part is not recursive, special model specifications are needed to address this feedback loop between the endogenous variables

Choice between SEM and GSEM

- Stata estimates SEM models through two sets of commands: Structural Equation Modeling (SEM) and Generalized Structural Equation Modeling (GSEM)
- SEM is used when all the endogenous variables are continuous and the model is at the single level
- GSEM is used when at least one endogenous variables is not continuous or when the SEM model includes multi-level, latent class, latent profile analysis
- Some estimation methods and modeling syntaxes are available only in SEM and others only in GSEM
- When choosing between SEM and GSEM, you need to balance between what SEM models to be examined and what results you want to generate by using SEM or GSEM commands

Estimation Methods

Stata has four estimation methods for SEM models:

- Two estimation methods are available for the SEM and GSEM commands
 - Maximum likelihood (ML)
 - Quasimaximum likelihood (QML)
- Two estimation methods are available for the SEM commands only
 - Asymptotic distribution free (ADF)
 - Maximum likelihood with missing values (MLMV)

Model Fit Statistics

1. Likelihood Ratio Chi-Square test:

$$x_{bs}^2 = 2\{\log L_s - \log L_b\}$$

$\log L_s$ is the log likelihood function of the saturated model and $\log L_b$ is the function of the fitted model. A value bigger than 0.05 indicates that the fitted model is not significantly different from the saturated model in explaining the data.

|

2. Akaike's Information Criterion:

$$AIC = -2 \log L_m + 2df_m$$

AIC is the log likelihood function of the model adjusted for the number of parameters estimated. Smaller AIC is better.

3. Swartz's Bayesian Information Criterion (BIC)

$$BIC = -2 \log L_m + Ndf_m$$

BIC is the log likelihood function of the model adjusted for the number of parameters estimated and the sample size. Smaller AIC is better.

Model Fit Statistics (Cont.)

4. Root Mean Square Error of Approximation (RMSEA)

$$RMSEA = \sqrt{\frac{(x_{ms}^2 - df_{ms})}{(N - 1)df_{ms}}}$$

RMSEA compares the current model to the saturated model by looking at the number of estimated parameters and the Chi-Square differences. A value of RMSEA smaller than 0.05 indicates a good fit, a value between 0.05 and 0.08 an adequate fit.

5. Tucker-Lewis Index (TLI)

$$TLI = \frac{\frac{X_{bs}^2}{df_{bs}} - \frac{x_{ms}^2}{df_{ms}}}{\frac{x_{bs}^2}{df_{bs}}}$$

TLI compares the current model with the baseline model by looking at the differences in the ratio of chi-square to the number of the relations between the current and baseline models. A value of TLI greater than 0.95 is good.

6. Comparative Fit Index

$$CFI = 1 - \frac{X_{ms}^2 - df_{ms}}{x_{bs}^2 - df_{bs}}$$

CFI compare the current model with the baseline model. A value of CFI greater than .95 is good.

Comparing Models

- When comparing nested models, use the differences in Likelihood Ratio test between models.
- When comparing non-nested models, use the differences in Akaike (AIC), Bayesian (BIC), or Sample-Size Adjusted BIC between models

Model Modification

Model fit can be improved by two ways:

- Add new paths to increase explanatory power of model. However, theoretical reasons need to be provided to justify the inclusion of these new paths in the model
- Reduce the number of parameters in the model
 - Delete nonsignificant paths from the model
 - Set the coefficients of paths to be equal
 - Fix the coefficients of paths to a certain value

Examples of Using Stata to Run Path Analysis

- See the handouts

BGSU



Center for Family and
Demographic Research

Strengths and Limitations of Using SEM in Fitting Path Models

Strengths:

- Ability to specify various models for different relations among variables, depending on theoretical frameworks
- Ability to distinguish among direct, indirect, and total effect of variables
- Comprehensive statistical tests on whether and how relations vary with other variables

Limitations:

- SEM techniques do not yield good results if the model is mis-specified or lack of theoretical basis
- Missing data and outliers influence the covariance and correlation matrices analyzed, which in turn influences SEM results. For models with categorical latent variables, GSEM still use listwise deletion and may not produce the best results
- A large sample size produces stable estimates of the covariance or correlation among variables, but it makes the model easier to be rejected
- There are multiple goodness-of-fit statistics to evaluate how well the proposed model fits the data. However, the cut-off criteria of these statistics are arbitrary and the conclusions from using different statistics may contradict each other
- There may be multiple equivalent models that fit data equally well

Conclusions

- SEM modeling allows for examining what direct and indirect effects a variable has on other variables, whether these effects vary with other variables, and whether a variable should be modeled as a latent variable
- Without SEM, a path model would be analyzed by estimating relations independently rather than simultaneously, likely producing inaccurate results
- Stata, SAS, LISREL, Amos, and Mplus all can analyze SEM models. Mplus is probably the most versatile one when it comes to analyze models involving categorical latent variable
- Fitting SEM models takes a lot of time because you need to modify the model one parameters at a time, so please allow yourself enough time for the analysis
- If you have any questions about SEM, feel free to contact Hsueh-Sheng Wu @372-3119 or wuh@bgsu.edu