Structural Equation Modeling in Brief (Part II): Path Diagrams and Coefficients

Colin R. Harbke
Department of Psychology
Western Illinois University

November 10, 2015
Center for Innovation in Teaching and Research (CITR) Workshop

Structural Equation Modeling

- SEM is an extremely general class of techniques, it includes:
  - Regression, Path analysis, Confirmatory factor analysis, and many others
  - In its most simple form, SEM is the combination of regression/ path analysis and factor analysis
- "SEM" typically refers to a model that evaluates structural processes between, at least one, latent variable and other variables
  - The *measurement model* is the confirmatory factor analysis portion
  - The *structural model* is the regression or path analysis portion
- The objective of structural equation models is to summarize the contents of a variance/covariance matrix
  - Through use of latent variables, *measurement model*
  - Through use of relationships, *structural model*
  - Correlation matrices are a special case of a variance-covariance matrix, where the variances have all be set to one (1)

Path Analysis in Amos

- Path analysis seeks to summarize a variance/covariance matrix only through relationships
  - In other words, it is only a *structural model*; there are no latent variables
- For example…

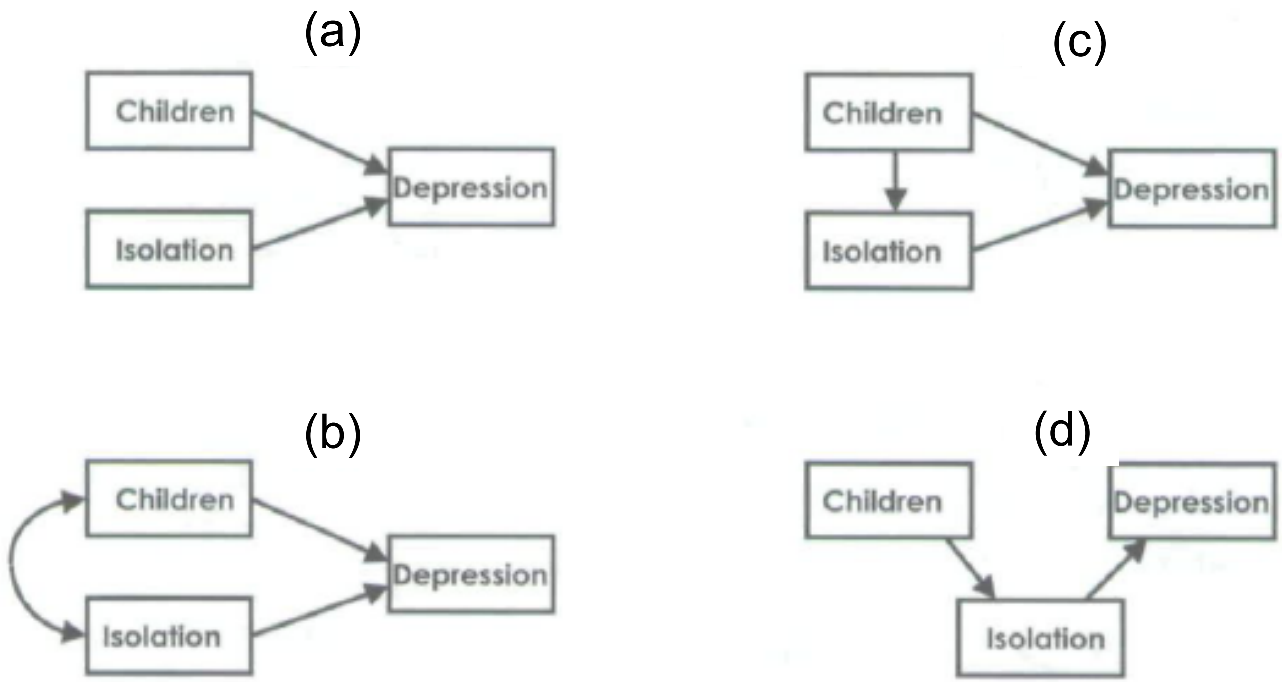
	Depression	Children	Isolation
Days	1046.25		
Children	18.38	22.88	
Isolation	72.46	8.75	94.26

Variance-Covariance Matrix
Raw Data Metric

	Depression	Children	Isolation
Depression	1.00		
Children	.12	1.00	
Isolation	.23	.19	1.00

Correlation Matrix
Standardized (*SD*) Metric

Basic Path Models



Path Analysis Variables and Structural Coefficients

- Because a single variable can serve as both an IV and a DV in path analysis, variables are described differently than for regression
  - *Exogenous variables* have directional arrows emerging from them and none pointing to them
    - They are like IVs in regression
    - May have error terms (with arrows from them) but often do not
  - *Endogenous variables* have directional arrows pointing to them and may also have arrows emerging from them
    - As for regression, unexplained DV variance is considered error
    - As such, endogenous variables require disturbance (or error) terms

- Structural coefficients are interpreted as regression weights; they describe the relationship between two variables after accounting for all other variables in the model

Input Matrix
Variance-Covariance Matrix

	Depression	Children	Isolation
Depression	1046.25		
Children	18.38	22.88	
Isolation	72.46	8.75	94.26

Summary is provided by estimating fewer parameters than there are elements in the variance covariance matrix.

The input matrix contains three variances; one for each variable.

The input matrix also contains three covariances.

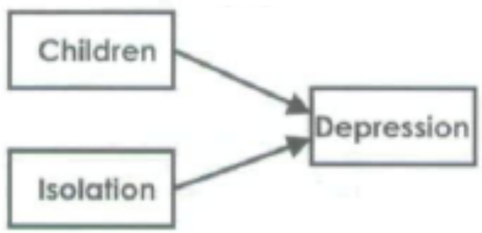
Six total elements/moments in the input matrix.

$$elements = \frac{p \times (p+1)}{2}$$

Degree of Summary

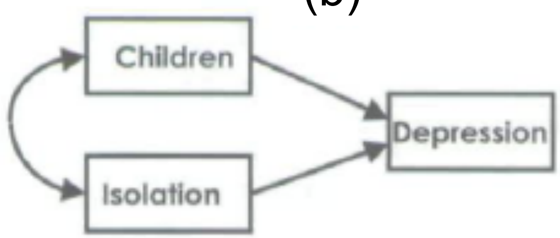
Typically each variance and path coefficient is estimated.

(a)



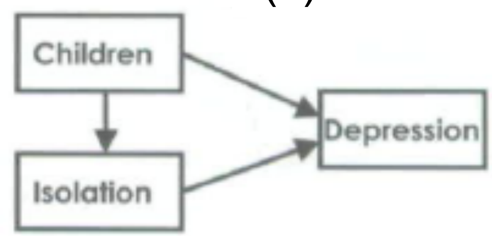
5 estimates

(b)



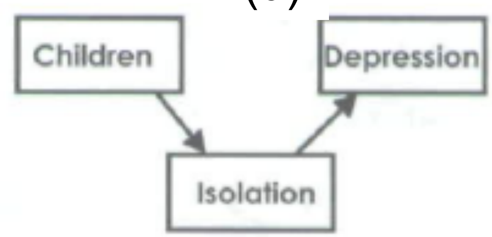
6 estimates

(c)



6 estimates;
0 degrees of freedom

(d)



5 estimates;
1 degree of freedom

$$df = elements - estimates$$

Model Degrees of Freedom

- Summary is provided by creating models with fewer estimated parameters than there are elements in the covariance matrix
  - *df* represent the number of independent data points (i.e., the count of elements that are neither fully constrained nor fully estimated)

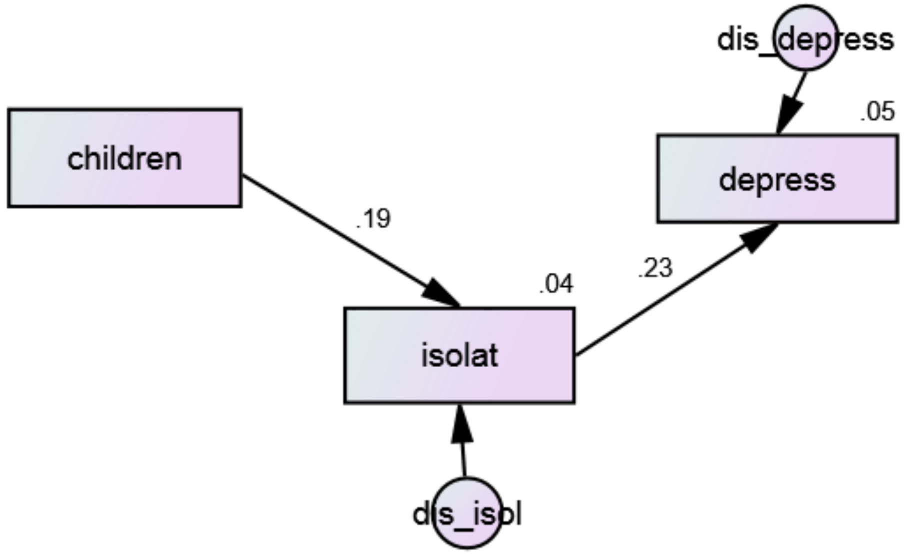
	children	isolat	depress
children	22.65		
isolat	8.66	93.31	
depress	18.19	71.73	1035.78

Input Matrix
Variance-Covariance Matrix

	children	isolat	depress
children	22.65		
isolat	8.66	93.31	
depress	6.65	71.73	1035.78

Model D Implied Matrix
Output Matrix

Model D provides an incomplete representation of the relationship between children and depression.



	children	isolat	depress
children	22.65		
isolat	8.66	93.31	
depress	18.19	71.73	1035.78

Input Matrix
Variance-Covariance Matrix

	children	isolat	depress
children	22.65		
isolat	8.66	93.31	
depress	6.65	71.73	1035.78

Model D Implied Matrix
Output Matrix

The “missing” or unestimated path between children and depression is the source of the incongruence between the input and output matrices.

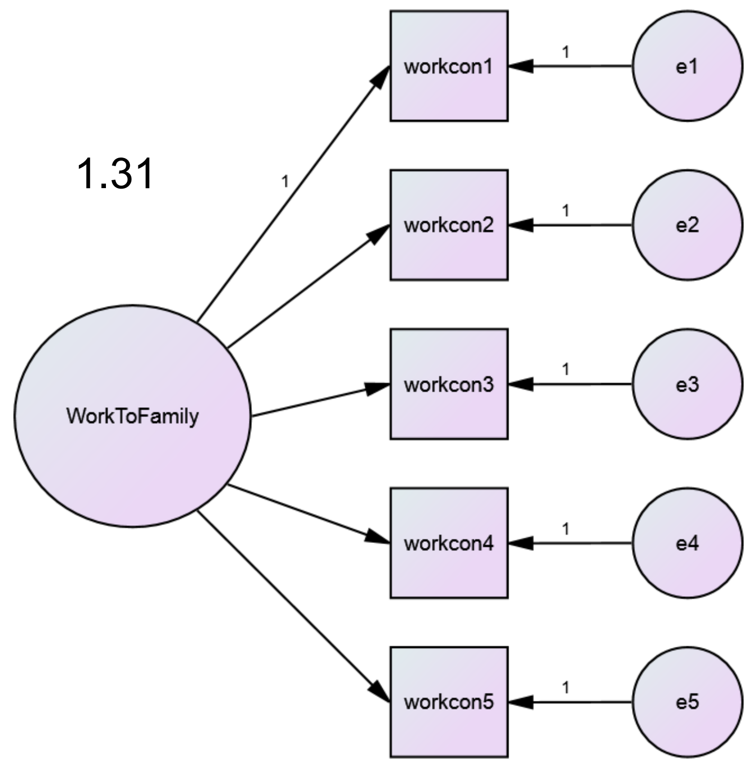
Path Analysis to SEM

- Path analysis is a type of SEM that describes relationships between observable variables
  - No latent variables are included and only the *structural model* is used to summarize the variance covariance matrix
- Including more variables in the analysis results in a larger input matrix and for more complex models
  - Rather than using scale scores, responses to individual items are used for the input matrix
  - Collections of items are summarized using latent variables or the *measurement model*
- "SEM" typically refers to a structural model examining relationships between, at least, one latent variable and others

Larger Input Matrix…

	Work5	Work4	Work3	Work2	Work1
Work5	1.25				
Work4	.65	.98			
Work3	.71	.69	1.12		
Work2	.64	.78	.73	1.00	
Work1	.82	.69	.81	.76	1.31

$$elements = \frac{p \times (p+1)}{2} = \frac{5 \times 6}{2} = 15$$



Larger Degrees of Freedom

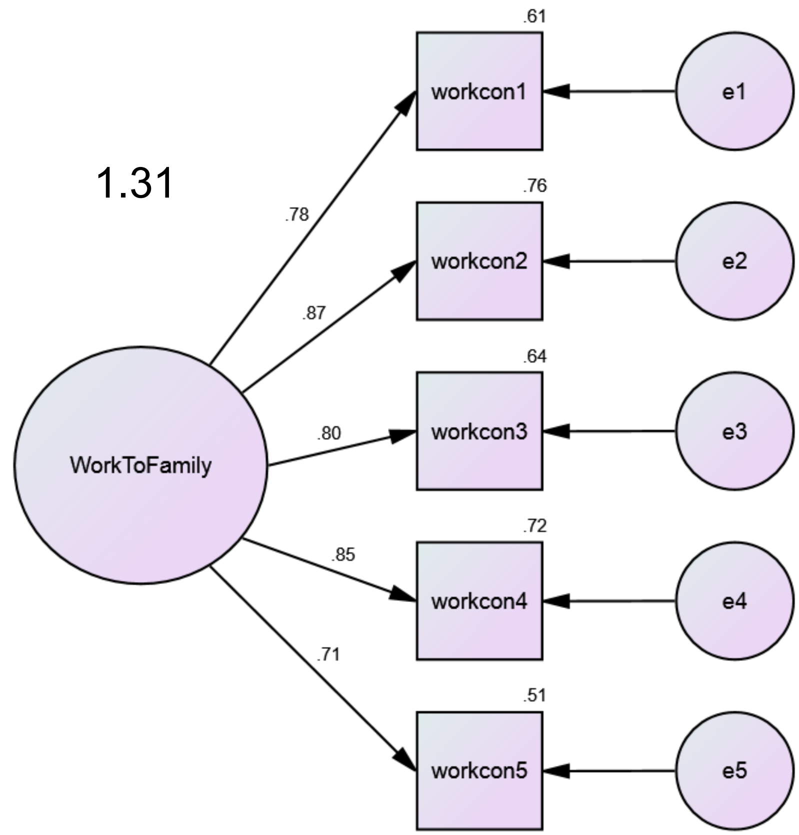
	Work5	Work4	Work3	Work2	Work1
Work5	1.25				
Work4	.65	.98			
Work3	.71	.69	1.12		
Work2	.64	.78	.73	1.00	
Work1	.82	.69	.81	.76	1.31

$$elements = \frac{p \times (p+1)}{2} = \frac{5 \times 6}{2} = 15$$

$$variances = 5\ observables + 1\ latent$$

$$covariances = 4\ factor\ loadings$$

$$df = 15 - (5+1) - 4 = 5$$

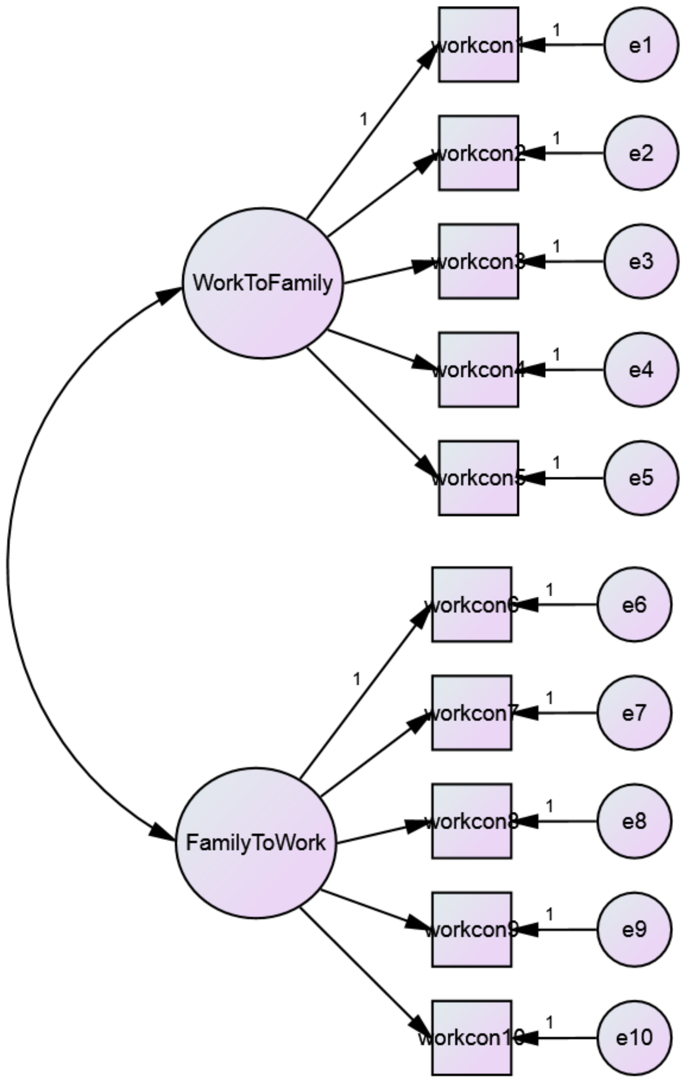


Even Larger Input Matrix

	Work5	Work4	Work3	Work2	Work1	Family10	Famly9	Family8	Family7	Family6
Work5	1.25									
Work4	.65	.98								
Work3	.71	.69	1.12							
Work2	.64	.78	.73	1.00						
Work1	.82	.69	.81	.76	1.31					
Family10	.30	.38	.31	.36	.38	.67				
Family9	.39	.38	.40	.37	.42	.44	.95			
Family8	.31	.36	.36	.35	.38	.41	.46	.69		
Family7	.36	.40	.40	.35	.43	.39	.51	.57	.81	
Family6	.73	.64	.62	.62	.71	.38	.44	.41	.48	1.00

$$elements = \frac{p \times (p+1)}{2} = \frac{10 \times 11}{2} = 55$$

Even Larger
Degrees of Freedom

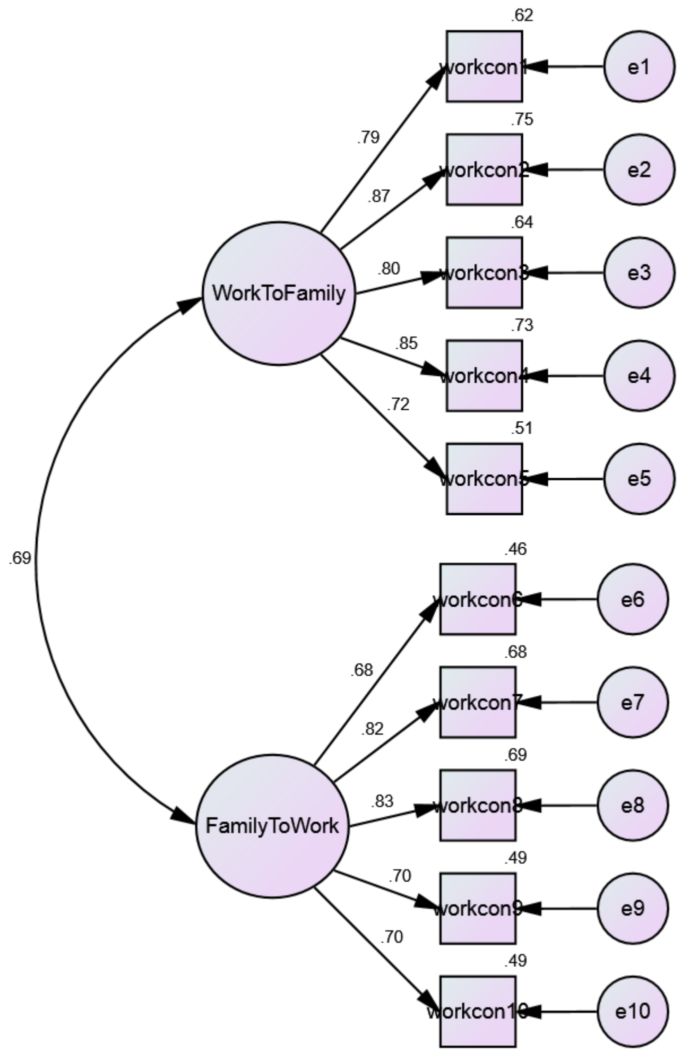


$$variances = 10\ observables + 2\ latents$$

$$covariances = 8\ factor\ loadings + 1\ interfactor$$

$$df = 55 - (10 + 2) - (8 + 1) = 34$$

Even Larger
Degrees of Freedom



$$variances = 10\ observables + 2\ latents$$

$$covariances = 8\ factor\ loadings + 1\ interfactor$$

$$df = 55 - (10 + 2) - (8 + 1) = 34$$

The *measurement model*, or confirmatory factor analysis portion, often provides the most degrees of freedom for a model.

Small Measurement Models

	Work3	Work2	Work1
Work3	1.12		
Work2	.73	1.00	
Work1	.81	.76	1.31

$$elements = \frac{p \times (p+1)}{2} = \frac{3 \times 4}{2} = 6$$

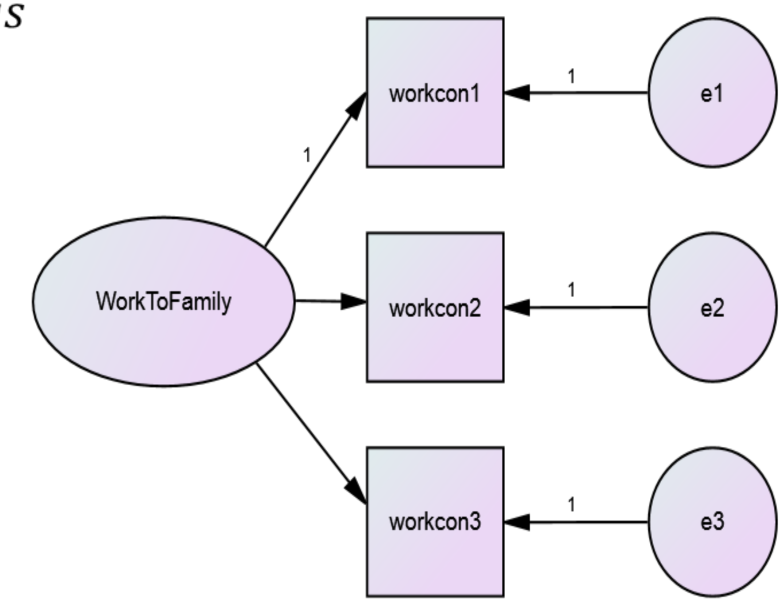
$$variances = 3\ observables + 1\ latent$$

$$covariances = 2\ factor\ loadings$$

$$df = 6 - (3+1) - (2) = 0$$

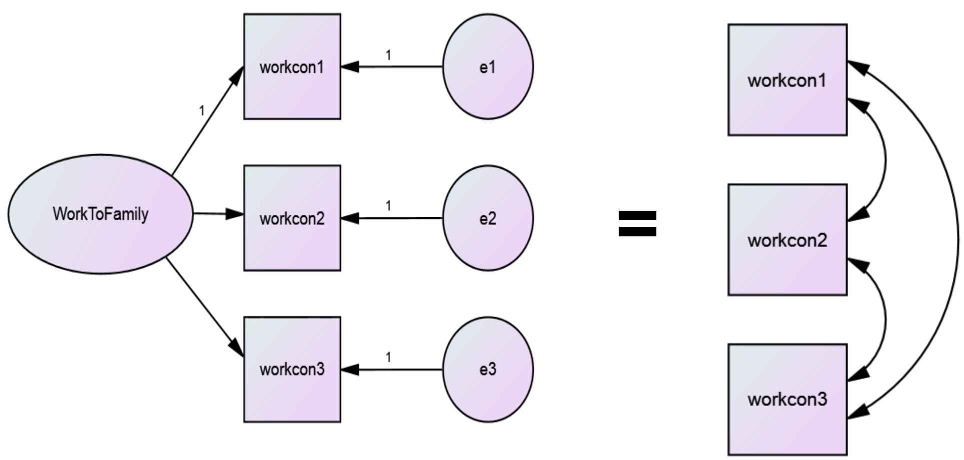
Each latent variable should be *identified*, in other words it should have positive degrees of freedom.

Without imposing additional constraints, each latent variable is just identified (*df* = 0) with three indicators. Thus, the typical minimum number of observables is three for each latent variable in the model.



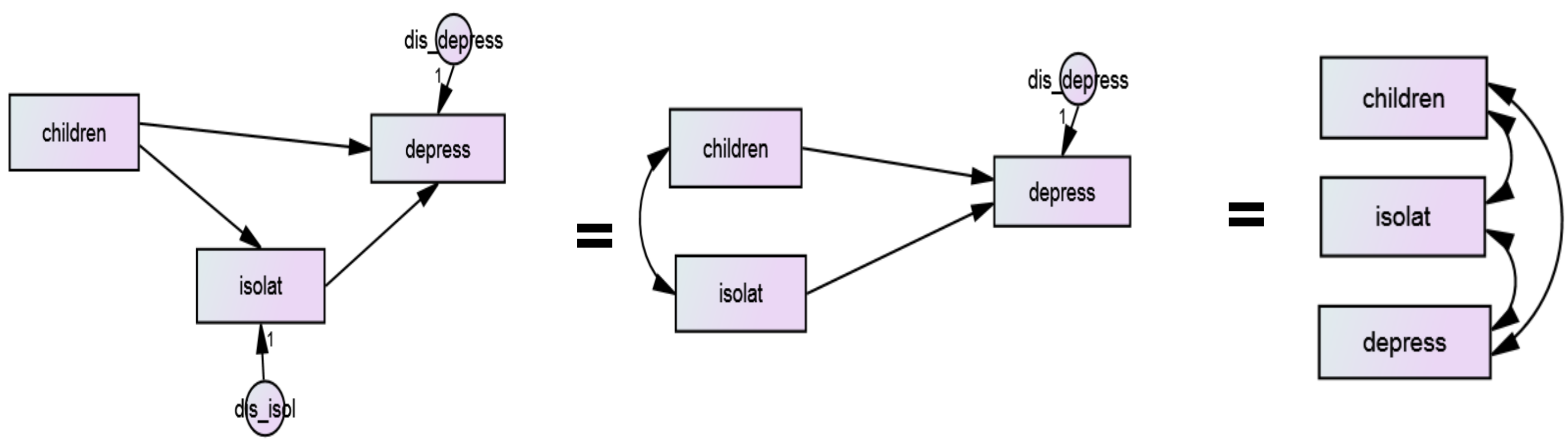
Alternative Representations with Equivalent Models

- Path diagrams provide a visual representation of the summarized (i.e., model implied) variance-covariance matrix
  - Path diagrams are readily interpreted
  - For any path diagram, however, alternative representations exist that suggest completely different interpretations
  - The degrees of freedom reflect the magnitude of the summary provided by the model/path diagram
  - When no summary is provided (*df* = 0), the alternative representations includes the variance covariance model itself



Alternative Representations with Equivalent Models

- Path diagrams provide a visual representation of the summarized (i.e., model implied) variance-covariance matrix
  - Two different models with the same degrees of freedom provide the same magnitude of summary (i.e., equivalent models). Thus, the models are mathematically identical, even though the path diagrams and interpretations may be **very** different
  - Also, be wary of causal inferences with correlational data



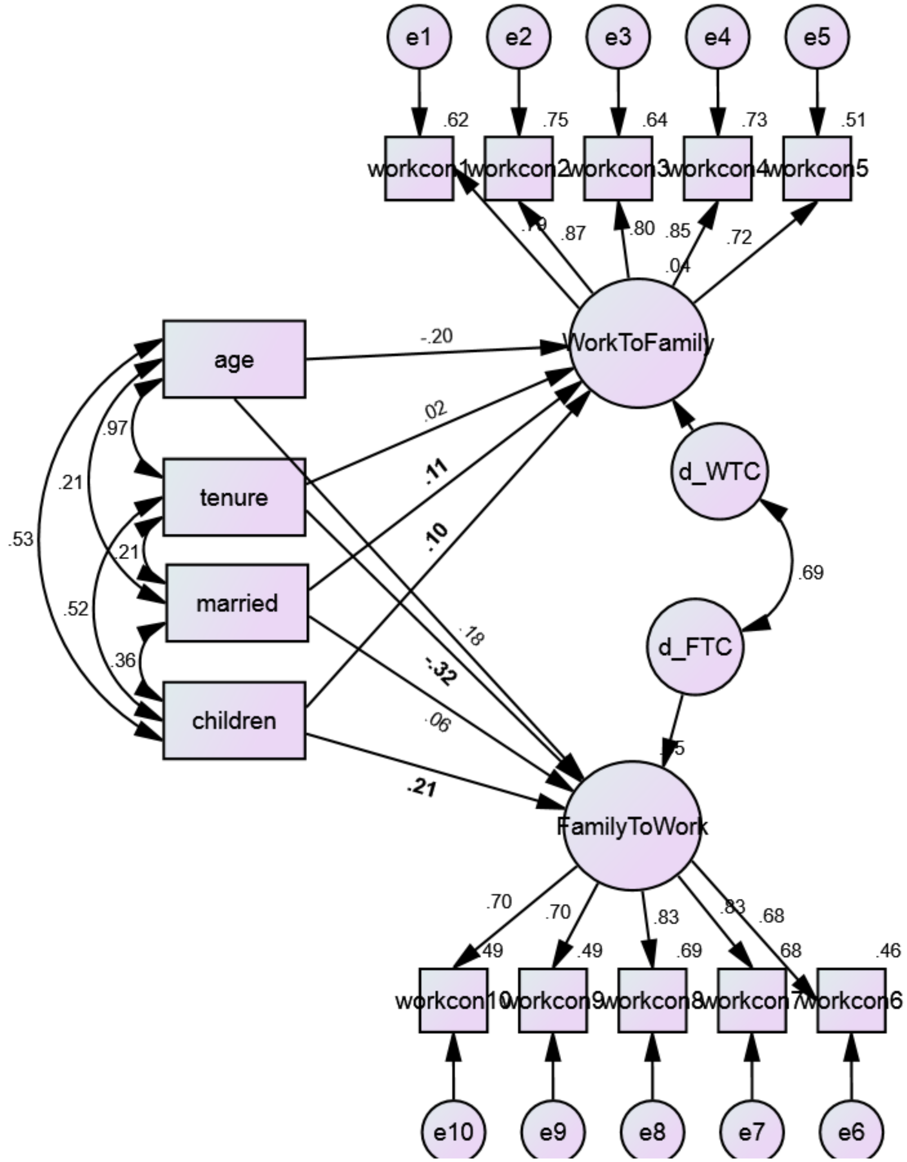
Path Diagrams

- Path diagrams are specified directly in Amos to summarize the input variance-covariance matrix
  - The path diagram reflects which variances and paths are estimated as part of the summary (i.e., unconstrained parameters)
  - Models with more degrees of freedom provide more summary
  - The *measurement model* (i.e., CFA portion) typically provides more degrees of freedom to the overall model than the *structural model* (i.e., path analysis portion)
- Each latent variable should have a minimum or three observable indicators
  - It is possible to use less than three, but additional constraints are needed to identify the latent variable
  - Larger variance-covariance matrices provide more information that can be summarized and, thus, often more degrees of freedom
- Be sure to consider alternative representations that suggest different interpretations, especially for models with low or zero *dfs*

Estimated Parameters and Coefficients

- Arrows or bi-directional arrows specified in Amos will produce an estimate of the relationship between variables when the data are fit to the model
  - Directional arrows between observables are regression weights
  - Bi-directional arrows between observables are covariances
  - Directional arrows from latent variables to observables are factor loadings
  - Bi-directional arrows between latent variables are inter-factor covariances
- Unstandardized and standardized coefficients are interpreted in the raw data metric or *SD* metric, respectively
  - Because the scaling is the same for are variables with standardized coefficients (i.e., correlations), they are often included on the output of path diagrams
  - Significance of estimated path coefficients can be determined by using critical ratios (*CR*) for the unstandardized estimates
- All coefficients are estimated after accounting for all other relationships/variances in the model (i.e., after partialling out everything else)
  - In short, all coefficients are regression weights -- not zero-order correlations

Predicting Work-to-Family and Family-to-Work Conflict



Structural model interpretation.

More tenure (in years) was associated with less Family-to-Work Conflict. Being married was associated with more Work-to-Family Conflict. Having children was associated with both Work-to-Family and Family-to-Work conflict. Despite high standardized coefficients for age with Work-to-Family and Family-to-Work Conflict, age was not significantly related to either, after accounting for other variables in the model.

Suppose we were to graph average SAT scores by the number of bathrooms a student has in his or her family home. That curve would also likely slope upward. (After all, people with more money buy larger homes with more bathrooms.) But it would be a mistake to conclude that installing an extra toilet raises yours kids' SAT scores..

-- Greg Mankiw

http://gregmankiw.blogspot.com/2009/08/least-surprising-correlation-of-all.html

Thank You!

Questions?

E-mail: CR-Harbke@wiu.edu

Remaining SEM in Brief Workshop

- Part III: Evaluation of Model Fit and Model Comparison will be presented Tuesday, 11/17 at 3:30pm
  - Inspection and comparison of overall model fit
  - Comparing competing models via difference tests
  - Summarizing results with path diagrams, tables, & text
  - http://www.wiu.edu/CITR/workshops/?action=showWS&id=2049

Assumptions for SEM

- Linearity
  - Relations among observables should be linearly related to one another
  - Linearity for relations with latent variables can be inferred observables for that latent variable are linearly related to other variables
  - Inspection of scatterplots to evaluate linearity and transformations can be used for non-linear relationships among variables

- Absence of multicolinearity or singularity
  - Very high multiple correlations among variables is problematic for SEM estimation techniques
  - Low tolerance (or determinant of the input matrix) will result in warnings or error messages or both
  - Removal of the redundant variables or converting them into combinations (e.g., parcels) will help

Assumptions for SEM

- Multivariate normality and absence of outliers
  - Many SEM estimation techniques assume normality
  - Outliers can be identified using stem-leaf, box-plots, etc.
    - Influential cases that are outliers are multiple variables are more of a concern
    - Mahalanobis distance is useful for detecting potentially influential cases
  - Mardia’s coefficient can be used to evaluate multivariate normality
  - Transformations can be used for univariate non-normal variables, or…
  - Robust estimation methods that incorporate corrections/scaling for non-normality can also be used
  - ADF (asymptotic distribution free) estimation does not assume normality, but can be inaccurate unless the sample size is very large (i.e., > 2500)

Assumptions for SEM

- Small residuals that are centered around zero and normally distributed
  - The residuals for SEM represent the discrepancy between the input covariance matrix and the model implied covariance matrix
  - Large residuals suggest that the model does not fit well (SRMR will be high)
  - A non-normal distribution of residuals suggests that the model does not fit well for particular elements in the variance covariance matrix
- If this is the case, modification indices can be inspected to learn where additional parameters can be added to incorporate the offending variables into the model (e.g., add cross-loadings, correlate errors)