

Neuendorf
Structural Equation Modeling

Structural equation modeling is useful in situations when we have a complicated set of relationships among variables as specified by theory. Two main methods have been employed to assess whether a complex and/or multi-step “causal” model is explained by the data at hand: (1) *Structural equation modeling* (SEM) solves multiple equations simultaneously, an improvement over (2) the older, by-hand process of *path analysis*. SEM allows for the combining of a structural/theoretic model with a measurement model. As Hair et al. note, SEM is “an extension of several multivariate techniques we have already studied, most notably multiple regression and factor analysis.”

Assumptions:

1. Measures are at the I/R level, independent observations, and distributions are normal and multivariate normal.
2. Even though SEM allows for multi-step models, theoretic constructs are deemed either “exogenous” (similar to independent) and “endogenous” (similar to dependent). Exogenous means that the construct is not predicted by any other construct; endogenous means there's at least one causal predictor of that variable (i.e., there is at least one causal path leading to it).
3. SEM allows for both (1) unmeasured, “latent” variables/constructs (structural/theoretic model only) and (2) measured, “observed” variables (measurement model or structural/theoretic model).
4. Relationships among the constructs and variables are linear.
4. Analysis is at the aggregate level. (e.g., In LISREL, the "data" consist of a correlation matrix--already aggregated over many respondents.)
6. The model and all its components are identified--generally, identification is related to the number of equations specified, and the number of coefficients to be estimated. [The attached page from Asher uses algebraic equations as a useful analogy.] The issue of model identification is a very complex one--you can't just visually peruse a model and declare it "identified." Some treatments of identification examine two “practical” conditions that test for model identification: The order condition and the rank condition (see attached pages from Maruyama). For a fuller discussion, see Asher's *Causal Modeling* (see attached), Blalock's *Theory Construction* or Heise's *Causal Analysis*.

Decisions to make:

1. Will there be a structural (theoretic) model? (YES--trick question; there has to be.) Will

there be a measurement model attached to this structural model? (Typically, yes--but this is OPTIONAL--this comes into play whenever there are two or more “observed” measures of a single “latent” construct)

2. What will the causal links be among the theoretic constructs? Any correlational links between exogenous variables? Any correlational links between errors of prediction on the endogenous variables? Any correlational links between measurement errors in the measurement model?
3. Will there be any nonrecursive "loops" in the model?
4. What estimation procedure will be used? LISREL offers maximum likelihood (default), two-stage least squares and several other options. AMOS offers maximum likelihood (default), generalized least squares and several other options.

Statistics:

1. Degrees of freedom = (roughly) # of elements in the correlation matrix - # of parameters to be estimated. It cannot be less than zero (if so, the model is not identified). Notice it is not related to n.
2. Path coefficients--are essentially like partial regression coefficients; show unique, standardized contribution of one variable to another's variance. For each coefficient, LISREL also provides a SE and a t-test to test whether it differs significantly from 0. AMOS provides a SE and a “C.R.” (critical ratio), which is Estimate/SE, and is equivalent to LISREL’s t. In both cases, anything greater than 1.96 is considered significant.
3. Goodness of fit (absolute, incremental, or parsimonious). Hair has a fine description of the choices on pp. 745-754. Arbuckle and Wothke have formulas and a classification scheme. Generally, these assess whether the model as a whole is "good" (well specified, fitting the data at hand). Some of the more commonly used goodness of fit measures are:
 - Chi-square (we want it to be small, nonsig.)
 - GFI (absolute goodness of fit index; we want it to be large, near 1)
 - AGFI (adjusted GFI; we want it to be large, over .90)
 - NFI (normed fit index; an incremental goodness of fit index; a recent “practical criterion of choice” (Byrne); we want it to be large, over .90)
 - CFI (comparative fit index; an incremental goodness of fit index; like the NFI with sample size taken into account; we want it to be large, over .95)
 - RMSEA (Root mean square error of approximation; avg. correlation among residuals; we want it to be small, substantially smaller than the original variable intercorrelations in the raw data matrix; Byrne presents standards of .05 or smaller as a good fit, .08 or smaller as a good fit, and .08-.10 as “mediocre fit”)

4. Multiple R²'s--one for each dependent variable. Indicates the proportion of the variance of that variable that is explained by the model.
5. Modification indices--for each unspecified potential link in a model, LISREL or AMOS calculates how much the Chi-square goodness of fit indicator will be reduced by including that link. If the incremental improvement is significant, you may wish to consider adding the link and rerunning.
6. IF you have a measurement model--"construct reliabilities" and "variance extracted" may be hand-calculated from loadings (relationships between measured variables and structural constructs). See Hair p. 777 for formulae.

References

Arbuckle, J. L. (2003). *AMOS 5.0 update to the AMOS user's guide*. Chicago, IL: SmallWaters Corp.

Arbuckle, J. L., & Wothke, W. (1995-99). *AMOS 4.0 user's guide*. Chicago, IL: SmallWaters Corp.

Asher, H. D. (1983). *Causal modeling* (2nd ed.). Newbury Park, CA: Sage Publications

Blalock, H. M. Jr. (Ed.). (1985). *Causal models in the social sciences* (2nd ed.). Chicago, IL: Aldine Publishing.

Byrne, B. M. (2001). *Structural equation modeling with AMOS: Basic concepts, applications, and programming*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL*. Baltimore, MD: The Johns Hopkins University Press.

Heise, D. R. (1975). *Causal analysis*. New York: Wiley.

Hoyle, R. H. (Ed.). (1995). *Structural equation modeling: Concepts, issues, and applications*. Thousand Oaks, CA: Sage Publications.

Maruyama, G. M. (1998). *Basics of structural equation modeling*. Thousand Oaks, CA: Sage Publications.

[Attachments: 3 pages from Asher, 2 pages from Maruyama]

From:

Asher, H. D. (1983). *Causal modeling* (2nd ed.). Newbury Park, CA: Sage Publications