STRUCTURAL EQUATION MODELS WITH LATENT VARIABLES

Douglas Baer
University of Victoria

Structural equation models (also referred to as "SEM models") have become very popular in the Social Sciences, especially in Psychology, Sociology, Education and some sub-disciplines of Business Administration. A major feature in the development of structural equation models from the earlier causal ("path") models of the 1960s and 1970s is the conceptualization of latent variables. The terms, "unmeasured variable models" and "latent variable models" refer to types of structural equation models that explicitly incorporate measurement error into the estimation of structural equation parameters, and treat observed ("manifest") variables as indicators of underlying constructs rather than perfectly measured representations of these same constructs. These models are quite general, and subsume many of the multivariate techniques typically dealt with in lower-level courses, including regression models, factor analysis, analysis of variance/analysis of covariance, principal components analysis, and path modeling. More recently, SEM models have provided an approach to the estimation of parameters in growth curve models for longitudinal data (these can also be estimated in the multilevel model framework).

In earlier usage, the models discussed in this course were, in the 1980s and early 1990s, often referred to as LISREL models. LISREL is now but one of the many computer programs now available to estimate SEM models; there are many others, including AMOS, which is now distributed with SPSS, and MPlus, both of which we will introduce during the course.

Early in the course, we will start with a scalar presentation of latent variable models and our class/lab examples will use the AMOS program, the SIMPLIS (scalar) version of LISREL, and scalar programming using MPlus. If there is any interest on the part of class participants, some examples using other software (SAS-CALIS; GLLAMM, EQS) will be presented.

After an introduction to SEM models in scalar terms, we will introduce the matrix-form representation of SEM models, emphasizing the use of LISREL software. Much of the literature (especially earlier literature) presents models using LISREL matrix notation. Next, we extend the models we have learned in two overlapping directions: multiple-group models and models for means and intercepts. Towards the end of the course, we shall cover some more advanced topics, including estimation in the presence of missing data and growth curve models for longitudinal data. These topics both require a thorough understanding of models for means and intercepts, which are usually covered in week 3.

What sort of a background is required for this course? At the very least, individuals should have taken the I.C.P.S.R. Regression Analysis II workshop or its equivalent (note that this is a second level graduate regression course), or its equivalent. A thorough familiarity with regression models is absolutely essential. Taking the two courses simultaneously (this course and the Regression Analysis II: Linear Models course) is not recommended. A good understanding of
the rudiments of matrix algebra is also important. While I.C.P.S.R. offers a set of Matrix Algebra Lectures early in the second session and while these lectures can help participants improve their matrix skills (indeed, beyond what is needed for the General Structural Equations course), the option of taking this course without any prior matrix algebra training should be considered only by those individuals who are not taking the course for formal university credit – and even at that, caution is appropriate. Some exposure to factor analysis will be helpful, since there are distinct parallels between some aspects of SEM modeling and factor analysis, but should not be considered essential.

Some form of an introduction to simultaneous equations and causal models is recommended but is not an absolute requirement. This year, the Simultaneous Equation Models workshop is offered during the second term; participants may wish to consider taking this course at the same time.

**Required and Recommended Readings:**

The major textbook for the class is a manuscript *An Introduction to Structural Equation Models for Latent Variables*, that I have prepared for the class. Chapters of this text will be available to I.C.P.S.R. participants for the cost of photocopying and should be considered essential for the course. (Instructions for obtaining these will be announced in the first class). It is **required**.

While this manuscript covers most of the material dealt with in the course, participants may wish to purchase copies of an additional text, since the ability to “triangulate” explanations is sometimes helpful in learning new techniques. Some copies of the following may be obtained at the bookstore (but participants could consider sharing a copy with a fellow participant or borrowing one of the multiple copies from the ICPSR library as required). These texts should be considered **recommended**.


2. David Kaplan, *Structural Equation Modeling*. Sage, 2000. This text is frustratingly terse at points and is thus not necessarily a good text for individuals who are not already familiar with latent variable structural equation models. It is in this sense better as a text to be read *after* a participant has finished the course (or at least most of the course). It contains useful treatments of some advanced topics: missing data, multilevel models, and latent growth curve models.

**Supplementary Readings:**

The following supplementary texts will also be useful during the course, but are not required:

1. An edited volume, called *Structural Equation Modeling: Concepts, Issues and Applications*, edited by Rick Hoyle (Sage, 1995). This text does not provide a thorough introduction to the area, but does contain some useful treatments of special topics and issues. Copies have not been ordered for the bookstore.

Some more advanced texts are more appropriate for the material covered in the last few days of class and for “further study” after a participant has taken the course. These include:


None of these more advanced texts has been ordered for the bookstore, but copies should be available in the ICPSR library.

In the past, we have sometimes ordered copies of software manuals. Because software vendors typically do not have liberal “returns” policies (if the bookstore does not sell copies, it cannot return them for a refund), we have not ordered any this year. Class participants should not require manuals for purposes of the course, since handouts on the use of the main software programs (AMOS, LISREL, MPlus) will be provided. And there will be multiple copies of the manuals both in the library and in the computer lab(s).

Instructions on the use of different computer software programs will be available in special class handouts dealing with each of the software programs we will be using in the course.

**Assignments and Exercises**

Most participants in this workshop do not attend for the purposes of obtaining formal course credit. For these individuals, it is still important to complete as many of the computer exercises as possible; without practical experience working with software and writing up “results,” participants are not likely to be able to conduct research of their own using the methods discussed in the course. There will be six computer exercises throughout the course.

For individuals taking the course for formal credit, grading will be based on:
1. Six computer exercises
2. A short (1-hour) test during week 2
3. A formal write up of research results
4. A 3-4 hour take-home test at the end of week 3

It is important that individuals who require a grade at the end of the course (taking the course for formal credit or would like ICPSR to write a letter indicating the grade that was received) identify themselves at the beginning of the course or mark “credit” or “grade” on their assignments. Assignments submitted by non-credit participants not requiring a grade are returned with comments and suggestions, but not with a grade.

Participants requiring a grade should ask for a copy of a “Grade Information” sheet which provides further information on the computation of formal grades for the course.

In the past, participants have asked if it would be possible to substitute any course requirements for a “major project” involving data that they are interested in working on. Unfortunately, the brevity of the summer program makes this alternative form impossible.

**Detailed Topic Outline and Reading List:**

The main reading for the course comes from the (Baer) manuscript to be distributed by I.C.P.S.R. (for the cost of photocopying), and from Schumacker and Lomax. Please pay careful attention to the distinction between readings labeled *Required*, *Recommended* or *Optional*, on one hand, and readings marked *Further Reading*. The first three types refer to readings that should be read as their contents are discussed in class (preferably before). Most of the articles and book chapters listed in this outline are, however, listed as *Further Reading*. These represent *continuations* of the material covered in class and will not necessarily be dealt with in class itself. In other words, it will not be necessary to read these materials before or immediately after the class covering the topic in question. Participants may, as time permits, read them when it is convenient.

Commonly used journals are referred to in short form, as follows:

“SEM” – Structural Equation Modeling
“SMR” – Sociological Methods and Research
“MBR” – Multivariate Behavioral Research
“PM” – Psychological Methods

*About the coverage of topics:*

Each topic area will normally take approximately two to three hours to cover, but some may take considerably more time and some may possibly take less.
1. An Overview

Required reading: Baer, chapter 1; Schumacker and Lomax, chapter 1.

Topics: Linear models for path/regression analysis; conceptualizing latent variables; structural equation models for latent variables; measurement error and its implications.

Optional: First chapter (Hoyle) in Hoyle Kaplan, chapter 1.

2. Covariance Algebra for Latent Variable Models

Required reading: Baer, chapter 2; Schumacker and Lomax chapters 2 & 3.

Optional: Kenny, Correlation and Causality (Wiley, 1979), chapter 3 (text available at ICPSR library)
Loehlin, chapter 1 (pp. 1-20; 26-31)

Topics: The basics; systems of equations; applications to path analysis models; reproduced covariances/correlations; direct and indirect effects; applications to factor models

3. Identification

Required reading: Baer, chapter 3; Schumaker and Lomax, chapter 6.

Optional: Kenny, chapter 8 pp. 134-138
Loehlin, pp. 82-83.
Kaplan, chapter 2 (to p. 24); also pp. 48-50*

*Kaplan's chapters all use matrix notation, which is not discussed until later

Topics: Under-identification and its implications; over-identification and its uses; establishing testable hypotheses; identification in factor models; identification in non-recursive causal models

Further reading:
Kenny, chapters 6, 7, 8.
Bollen, chapter 7
*more advanced; may not necessarily be discussed in class or may be discussed very briefly

4. Scalar Programming and Model Specification

Required Reading: Sections of Baer, chapter 6: (a), (b)

A special handout will be provided

This discussion will emphasize the use of AMOS with both SPSS system files and raw covariance matrices. If time permits, some attention will be given to the "mechanical" issues of file construction (how to construct appropriate raw data and covariance matrices from SPSS and/or SAS, etc.).

Computer Exercise #1 (due Monday, July 25)
This exercise will use AMOS.

A small amount of time will be devoted to the use of the following packages for scalar model notation:

- LISREL (Simplis interface)
- MPlus

The following will be covered fairly tersely and only if there is class interest:

- The CALIS procedure in SAS
- EQS

[Refer to Schumaker and Lomax, chapter 8]

The amount of time, if any, spent on each of the above will depend upon the extent of interest shown by class participants. It will not be possible to provide details instructions and examples, though if there is enough interest, special extra lab demonstrations can be arranged.

5. Matrix Algebra for Path and Factor Models and LISREL model notation

Required Reading: Baer, chapter 7

Optional Reading:
Bollen, Appendix A and chapter 2 or

Topics: covariance structure algebra in matrix terms; reproduced covariances in matrix terms. LISREL model notation for structural equations involving observed variables; LISREL model notation for confirmatory factor models; LISREL model notation for structural equations involving latent variables; the exogenous-endogenous distinction; comparisons with scalar models

6. Programming LISREL

*A special handout will be provided.*

Topics: model specification; constraints; the use of PRELIS; multiple group problems

*Computer Exercise #2: LISREL programming (matrix form), due Wednesday July 30*

7. Scaling and Interpretation Issues

Required reading: Baer, chapter 4

*Optional:* Schumaker and Lomax, chapters 4 & 5.
Kaplan, pp. 34-39*

*explanation is in matrix terms*

Topics: establishing a metric for latent variables; constructing linear composites; fixed and free parameters in models; covariances among latent variables; variances of latent variables; standardized solutions; mixing (single-indicator) manifest and (multiple-indicator) latent variables

*Further Reading:*

W. Bielby, Arbitrary Metrics in Multiple Indicator Models, SMR, 15(1), 1986, pp. 3-23
Bollen, chapter 8, pp. 349-355.
R. MacCallum, Model Specification, chapter 2 in Hoyle.

8. Estimation, Model Fit and Model Improvement

Required reading: Baer, chapter 5

*Recommended:* Schumaker and Lomax, chapter 7.

*Optional:* Loehlin, chapter 2; chapter 7 (pp. 195-204)
Kaplan, chapter 2, pp. 24-34; chapter 6.
Topics: Reproduced vs. empirical covariance matrices; the chi-square test for model fit; incremental chi-square tests; some goodness of fit indices; Lagrange Multiplier tests, modification indices, Wald tests; testable and non-testable hypotheses; exploratory modification of models.

Further reading:


9. Fit Indices: Uses and Abuses

Recommended:


Topics: Sample size bias; parsimony; typical values in sparse vs. dense models; relationship to model replication issues.

Further Reading:
L. Hu and P. Bentler, Evaluating Model Fit, chapter 5 in Hoyle.
Ding, L., Velicer, W., & Harlow, L. Effects of Estimation Methods, Number of Indicators Per Factor, and Improper Solutions on Structural Equation Modeling Fit Indices., SEM, 2(2), 1995, 119-144.
Ken Bollen, A New Incremental Fit Index, SMR, 17(3), 1989, pp. 303-316.


Jackson, Dennis, “Revisiting Sample Size and Number of Parameter Estimates: Some Support for the N:q Hypothesis,” SEM, 10(1), 2003, 128-141.


Assignment: Write-up of research results. Due Thursday, August 7

10. Problems and Issues

Optional: Bollen, chapter 7, pp. 281-286.
Kaplan, pp. 79-80.

Topics: Sample size and goodness of fit; improper parameter estimates; collinearity; missing data; identification in complex models; equality constraints; inequality constraints; categorical exogenous variables; weighting and stratified samples [if time permits: power in significance tests #] 

Further reading:

Bollen, chapter 8, esp. pp. 338-355.
Ridskopf, David Parameterizing Inequality Constraints on Unique Variances Psychometrika, 48, 1983, 73-83.
Jackson, D. The Effect of the Number of Observations per Parameter in Misspecified Confirmatory Facotr Analytic Models, SEM, 14(1), 2007, 48-76.

11. Simultaneous Analysis in Multiple Groups

Required Reading: Baer, chapter 8

Optional: Bollen, chapter 8, pp. 355-365
Schumaker and Lomax, chapter 10.3.
Kaplan, chapter 4

Topics: Introduction: replicating models across groups; across-group parameter constraints; testing for measurement equivalency; testing for equivalency of causal (structural equation) effects; comparisons with analysis of covariance designs

Further Reading:


Computer exercise #3: Multiple Group Models Due: Monday, August 4

12. Comparisons involving Factor Means

Required Reading: Baer, chapter 9.

Optional Reading: Loehlin, pp. 204-210.
Kaplan, pp. 68-70.

Topics: Adding intercepts to latent variable models; factor mean comparisons; mean comparisons in structural equation models

Further Reading:
Bollen, chapter 7, pp. 306-311
Bollen, chapter 8, pp. 365-368
*more advanced treatment of topic to be covered only if time permits

Computer exercise #4: Multiple Group Models for means and intercepts
Due: Friday, August 8

13. Distributional Assumptions, the ADF Estimator, Robust Test Statistics, Bootstrapping

Recommended: Bollen, pp. 415-432
Kaplan, chapter 5 (to page 87)

Topics: Data screening; data transformations for continuous data; discrete and coarsely-categorized variables; dichotomous variables (as X-variables; as indicators); robustness of ML estimator; “robust” statistics; the ADF estimator; polychoric correlations for ordinal data

Further Reading:
Browne, M. Asymptotically Distribution-Free Methods for the Analysis of Covariance Structures, British Journal of Mathematical and Statistical Psychology, 37, 1984, pp. 62-83. (mathematically intense and will not be discussed in full detail, but provides the basis for what has come to be known as the ADF estimator)
S. West, J. Finch and P. Curran, Structural Equation Models with Nonnormal Variables: Problems and Remedies, chapter 4 in Hoyle.
S. Green et al, Effect of the Number of Scale Points on Chi-Square Fit Indices in Confirmatory Factor Analysis, SEM, 4(2), 1997, 108-120.(see also Yung and Bentler under "Bootstrapping Approaches")

Further Reading on Bootstrapping:
Yung, Y.F. and P. Bentler, Bootstrapping Techniques in the Analysis of Mean and Covariance Structures. In M & S.

Further Reading on Sample Size and Small Samples:

14. Missing Data in SEM Models

Required:
Bollen, pp. 369-376.

Recommended: Kaplan, chapter 5, pp. 87-96.
Further Reading:

J. Arbuckle, Full Information Estimation in the Presence of Incomplete Data, chapter 10 in M & S.
*Gold, Michael, Peter Bentler and Kevin Kim, “A Comparison of Maximum-Likelihood and Asymptotically Distribution-Free Methods of Treating Incomplete Nonnormal Data,” SEM, 10(1), 2003, 47-79
*Graham, John, “Adding Missing-Data-Relevant Variables to FIML-Based Structural Equation Models,” SEM, 10(1), 2003, 80-100.

Computer exercise #5: A model for “messy” data (non-normal with missing cases). Due: Tuesday, August 12

15. Item Parcelling

Further Reading:


16. Models for Panel Data

Topics: Direct and indirect effects in causal models; non-recursive models contemporaneous and lagged effects in a two-wave panel model; correlated measurement error in panel models; relationship between SEM approaches and ARMA time-series models See also topic 17, which is related, below.


Further Reading:

Karl Jöreskog and Dag Sörbom, Advances in Factor Analysis and Structural Equation Models, chapters 5 (Statistical Methods for Analysis of Longitudinal Data) and 6 (Detection of Correlated Errors in Longitudinal Data).


Lawrence Mayer and Steven Carrol, Testing for Lagged, Cotemporal and Total Dependence in Cross-Lagged Panel Analysis, SMR, 16(2), 1987, pp. 187-217


17. Growth Curve Models

Further Reading:

Computer Exercise #6: A Longitudinal Data Model. Due: Thursday, August 14.

One or more of the following topics might be covered if time permits:

18. Polynomials, Interactions and Non-Linear Models ##

Read: Bollen, chapter 9, pp. 403-415
L. Hayduk, Structural Equation Modeling with LISREL, chapter 7, pp. 219-244.

Further Reading:

R. Schumacker and G. Marcoulides (eds.) Interaction and Nonlinear. (Entire text, but especially chapters by Rigdon, Schumacker and Wotchke [ch. 1], Jonsson [ch. 2], Bollen and Paxton [ch. 6], Joreskog [ch. 11]

19. Latent Class Analysis; Mixture Models ###
Further Reading:


B. Muthen, Goodness of Fit with Categorical and Other Nonnormal Variables, chapter 9 in Bollen and Long.


Allan McCutcheon, Latent Class Analysis. Sage Quantitative Applications Paper No. 64 (1987)


20 Alternative Estimation Methods ###

Further Reading:


K. Bollen. A Limited Information Estimator for LISREL Models With or Without Heteroscedastic Errors. In M & S.


21. Multilevel Analysis ###

*Further Reading:*


Patrick Curran “Have Multilevel Models Been Structural Equation Models All Along?” *MBR*, 39(4), 2003, 529-569