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, more recently, in Business and Public Administration and various applied health sciences (e.g., Nursing). 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), and for an approach to the problem of the unbiased estimation of parameters in the presence of missing data.

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 MPlus, AMOS, EQS, a set of SEM procedures distributed with R, and a SAS procedure. In this course, we will emphasize MPlus, which is the most advanced of the SEM packages currently available and fast becoming the most popular. AMOS, which is distributed by SPSS and which is probably the second most popular SEM package in part because of its fairly intuitive interface, its ability to draw nice diagrams and the fact that many universities have SPSS site licenses, will not play a prominent role in the course, but those who are interested in learning how to use it can take part in a couple of optional “labs” outside of the regular class period (probably in the early evening). Participants who cannot attend these optional labs should still be able to follow the self-instructional guides that will be handed out (and thus teach themselves how to use the software). If there is interest on the part of participants, opportunities (self-instructional guides, a possible extra optional “lab” outside of regular class time) will be provided to learn how to program other SEM software.

After an introduction to SEM models in scalar terms, and an introduction to the extension of our models to simultaneous multiple-group estimation, we will briefly introduce the matrix-form representation of SEM models. Much of the early literature literature presents models using LISREL matrix notation (though this is no longer the norm), and some software (e.g., LISREL) is programmed primarily in matrix form. Next, we extend the models we have learned to models for means and intercepts, which are useful both for factor mean comparisons in the multiple-group case and for longitudinal data analysis in the single (or multiple) group case. 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. If time permits, we will cover extensions to models with ordinal or binary indicators but continuous latent variables, and to models with categorical latent variables.

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.

Two I.C.P.S.R. courses which are available at different times in the summer of 2011 will complement the material covered in this course. The simultaneous equation models course covers many details in causal modeling which are dealt with only in cursory fashion in the structural equation models course, and the longitudinal analysis course will be of interest to those intending to analyze panel data (which we cover from a different perspective in the last week and a half of the structural equation models course).

**Required and Recommended Readings:**

The major readings for this course can be found in a series of 15-30 page written course notes which will be posted on ICPSR “local” servers (in the Helen Newberry building) and on the “CTools” pages available via the web. These are exclusively for the use of ICPSR participants and are in PDF format, and can be downloaded. If there is interest, arrangements can be made to have these printed off and made available to class members for the cost of printing, but in recent years participants have found it more useful to simply download the material in electronic form.

While the course notes cover 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. The most useful of these will be:


This text is available at Ulrich’s bookstore. A secondary text will be:

The following texts cover some of the more advanced topics we deal with. A small number of copies has been ordered for the bookstore. These should not be considered essential:


The following texts have not been ordered for the bookstore, but multiple copies should be available for borrowing at the ICPSR library in Helen Newberry. Again, these should not be considered essential:


Manuals for MPlus software should be available in the computer labs in Helen Newberry, but a PDF of the entire manual should be freely available from the MPlus web site (www.statmodel.com). A special class handout will also be available in hard copy and PDF format.

**Assignments and Exercises**

Most participants in this workshop do not attend for the purposes of obtaining formal course credit. For non-credit participants, it is 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 five computer exercises throughout the course and three exercises not involving the use of computer software. Individuals taking the course for credit (or to receive a letter with a grade) will also be asked to write an exam at the end of the course. At the same time as the test is conducted, a special “lab” will be available for other participants.

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.
Topic Outline:

Week 1
- Basics: manifest vs. latent variables; measurement error and its implications
- Hypothesis testing; covariance algebra for structural equation models; MPlus software basics
- Estimation; identification; interpreting results
- Scaling and interpretation issues; constraints in SEM models; variances of latent variables; model fit and model improvement
- Problems and issues (convergence, negative variance estimates); start values; higher-order latent variables (“2nd order factor models”); the use of “phantom variables”

Week 2
- General linear parameter constraints; Multiple-group models (the 2-group case)
- Multiple group models: the k group case; measurement equivalence, construct equation equivalence, other tests
- LISREL matrix notation for structural equations
- Model assumptions; performance of ML estimator under non-normality; transformations; writing up results from structural equation models
- Alternative parameterizations; reproduced covariances in matrix terms; goodness of fit measures
- Models for means and intercepts

Week 3
- Application of mean/intercept models to multi-group problems
- Parallel and non-parallel slope models
- Approaches to missing data
- Ordinal data, non-normal data: alternative estimators, scaled test statistics, bootstrapping
- Longitudinal data in SEM models

Week 4
- Growth curve models for single-indicator variables
- Growth curve models for multiple-indicator variables
- Mixture models for categorical outcomes
- General approaches to categorical data (latent class models, variants).