Psyc 8501-001
Introduction to Structural Equation Modeling

Course Syllabus

Meeting Time: Mondays, 9:00 — 11:30
Room: 1023 Millmont, room 123

Instructor: Steven M. Boker
Office: Millmont 115
Office Hours: Thursday 9:00–11:30 or by appointment
Email: boker@virginia.edu
OpenMx Web Page: http://openmx.psyc.virginia.edu/

1 Focus

Structural Equation Modeling (SEM) is a statistical technique to estimate regression coefficients and assess overall fit of a theory to data that involve multivariate outcomes. The covariances between variables can be decomposed to assess their "structure", that is to say, whether a parsimonious set of relations between the measured variables can account for the observed covariances.

SEM is typically used in settings where there are many individuals measured on a set of variables or when each individual is measured many times. Thus one typically needs on the order of a hundred rows in one's data matrix before SEM begins to give stable parameter estimates.

SEM is a generalization of other modeling techniques such as the General Linear Model and can be used to represent theories that involve specific constraints on regression parameters. Because of this, SEM is used in large epidemiological studies to compare hypotheses that involve specific etiologies. For instance, behavioral genetics studies can use twin designs that distinguish between the genetic similarity of monozygotic and dyzygotic twins. Or longitudinal studies may wish to understand differences in growth curve trajectories between multiple groups.

SEM excels in a number of other areas, such as accounting for missing data when the data are not conditioned on an outcome variable, measurement models where a small number of factors are hypothesized to account for a large number of variables, and time–delay models where change over time is predicted as a function of either intrinsic or extrinsic dynamics. SEM models are also used to test hypotheses involving mediating and moderating variables.

The first part of the course will introduce the rationale and scope of SEM and cover the background required to understand how the models work. We will begin by reviewing the matrix algebra required to read and understand articles that use SEM and to create one's own SEM models. Next, we will cover how the expected covariance from models is derived so that we gain an understanding of how regression relationships between pairs of variables account for covariance among many variables. We will work through these ideas both using matrix algebra and path diagrams so that the correspondence between graphical and mathematical representations of SEM can be used as a framework to quickly understand and communicate complicated model structures.

The second part of the class will cover a range of practical applications of SEM. Students will be asked to contribute data along with alternate theories for their data to be used as examples. We will work through modeling these use–cases in class. The exact examples that are covered will depend on the interests and research of the students. However, if an important model class is not represented in students' data, an example will be contributed by the instructor.

Computer work associated with the course will primarily involve the SEM software OpenMx which runs as part of R. OpenMx is free for download and runs on PCs, Macs, and Linux. We will also use scripts in R for preprocessing data and for graphical diagnostics. Students will also acquire sufficient skill in writing OpenMx scripts such that they can perform SEM in practical applications on their own research problems.

2 Prerequisites

Familiarity with regression, covariance, the general linear model, and exploratory factor analysis is assumed. We will review some of the necessary concepts from matrix algebra and calculation of the covariance of linear combinations, but if the student has not been exposed to these before, the student will need to spend extra time outside of class in order to obtain the most from the remainder of the semester.

3 Student Evaluation

Evaluation and course grades will be based on an article written in APA style as if for publication. There is no page limit, but you must be first author, it must be new work, and it must include an SEM analysis comparing at least two models. Hopefully you will actually submit it for publication! The first half of the written article will be due in class on March 15, comprising the introduction and methods section of the article. The final version of the written article will be due on Monday, May 10 by 11:30 am.

The students who obtain the most from this class are likely to be those who have data that is amenable to analysis by SEM and use the class as a way of obtaining guidance while writing up an SEM–based article for publication. By doing so, (a) it will further the student's progress towards degree and a job in the short term and (b) the concepts and practices learned in the class are much more likely to be integrated into the students' long term career trajectory. The wise student begins thinking in the very first class about what they want to know from their data and what alternative explanations might be tested. As we progress through the class, formalized models for those explanations are then more likely to be incorporated into the student's thinking.

Thus, there are two types of student evaluation here. The instructor's evaluation of the students is a given for those taking the class for a grade. But more important is the students' evaluation of the concepts and models in order to incorporate them into immediate practical use.

4 Textbooks and Handouts

Two textbooks are recommended. The first is *Basics of Structural Equation Modeling* by Geoffrey Maruyama. This book covers the essential ideas in a readable manner and may help your understanding when read as an adjunct to the lectures. The second book that you may want to have on your shelf is Raykov & Marcolides' *A First Course in Structural Equation Modeling.* This book has examples from EQS, Lisrel, and MPlus, so it may be of particular use if you decide that you would rather use one of those programs. If you were to just buy one book, I would recommend Maruyama.

In addition, the OpenMx manual is available at `http://openmx.virginia.edu`. There are discussion forums on the website that we will use to talk about models and examples from class.

As a reference to the use of R, I recommend Peter Dalgaard's *Introductory Statistics with R.* I make this recommendation based on feedback from students who have found it to be the most helpful R textbook that I have recommended over the years.

Finally, there will be supplemental journal article readings that are either methodological advances or examples of practical applications.

5 Timeline

Date	Description	Readings
Jan. 25	Overview of SEM: What is it, when should I use it, and where did it come from?	(Maruyama, 1998, Chapters 1 & 2)
Feb. 1	Review of the basics of matrix algebra and the covariance of linear combinations.	Handouts.
Feb. 8	Path analysis, path diagrams, and components of covariance: The model expectations.	(Boker & McArdle, 2005) (Boker, McArdle, & Neale, 2002) (McArdle & Boker, 1990) (Maruyama, 1998, Chapter 3)
Feb. 15	Fitting manifest variable structural models with Mx.	Handouts, (Neale, Boker, Xie, & Maes, 2003, Chapters 2 & 3), or as an alternative (Rakov & Marcoulides, 2006, Chapters 2 & 3)
Feb. 22	Fit functions and model comparison.	(Maruyama, 1998, Chapter 10) (Neale et al., 2003, Chapter 5.3)
Mar. 1	Confirmatory factor analysis (measurement models) and latent variables.	(Maruyama, 1998, Chapter 7) or (Rakov & Marcoulides, 2006, Chapters 4) (McArdle, 1990)
Mar. 8	Spring Break	
Mar. 15	First half of paper is due at 11:30 am.	
Mar. 15	Latent variable structural models.	(Maruyama, 1998, Chapter 8) or (Rakov & Marcoulides, 2006, Chapters 5)
Mar. 22	Mediation models.	(Cole & Maxwell, 2003) (Baron & Kenny, 1986)
Mar. 29	Full information maximum likelihood: missing data, unbalanced data, and moderators.	Handout, (Neale et al., 2003, Chapters 5.2 & 6.4) (Graham, Taylor, Olchowski, & Cumsille, 2006) (McArdle, 1994)
Apr. 5	Longitudinal latent change / latent difference models.	(McArdle & Hamagami, 2001)
Apr. 12	Multigroup and behavior genetic models.	(Neale, Boker, Bergeman, & Maes, 2006)
Apr. 19	Growth curve models.	(Duncan & Duncan, 1995) (McArdle & Epstein, 1987)
Apr. 26	Latent Differential Equations.	(Boker, Neale, & Rausch, 2004)
May 3	Multilevel Models.	
May 10	Final Paper is due at 11:30 am.	

References

Baron, R. M., & Kenny, D. A. (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182.

Boker, S. M., & McArdle, J. J. (2005). Path analysis and path diagrams. In B. Everitt & D. Howell (Eds.), *Encyclopedia of statistics in behavioral science (vol. 3)* (pp. 1529–1531). New York: John Wiley & Sons.

Boker, S. M., McArdle, J. J., & Neale, M. C. (2002). An algorithm for the hierarchical organization of path diagrams and calculation of components of covariance between variables. *Structural Equation Modeling*, *9*(2), 174–194.

Boker, S. M., Neale, M. C., & Rausch, J. (2004). Latent differential equation modeling with multivariate multi-occasion indicators. In K. van Montfort, H. Oud, & A. Satorra (Eds.), *Recent developments on structural equation models: Theory and applications* (pp. 151–174). Dordrecht, Netherlands: Kluwer Academic Publishers.

Cole, D. A., & Maxwell, S. E. (2003). Testing mediational models with longitudinal data: Questions and tips in the use of structural equation modeling. *Journal of Abnormal Psychology*, *29*(4), 409–454.

Duncan, S. C., & Duncan, T. E. (1995). Modeling the processes of development via latent variable growth curve methodology. *Structural equation modeling*, *2*, 187–213.

Graham, J. W., Taylor, B. J., Olchowski, A. E., & Cumsille, P. E. (2006). Planned missing data designs in psychological research. *Psychological Methods*, *11*(4), 323–343.

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

McArdle, J. J. (1990). Principles versus principals of structural factor analyses. *Multivariate Behavioral Research*, *25*(1), 81–87.

McArdle, J. J. (1994). Structural factor analysis experiments with incomplete data. *Multivariate Behavioral Research*, *29*(4), 409–454.

McArdle, J. J., & Boker, S. M. (1990). *Rampath.* Hillsdale, NJ: Lawrence Erlbaum.

McArdle, J. J., & Epstein, D. (1987). Latent growth curves within developmental structural equation models. *Child Development*, *58*, 110–133.

McArdle, J. J., & Hamagami, F. (2001). Linear dynamic analyses of incomplete longitudinal data. In L. Collins & A. Sayer (Eds.), *New methods for the analysis of change* (pp. 137–176). Washington, DC: American Psychological Association.

Neale, M. C., Boker, S. M., Bergeman, C. S., & Maes, H. M. (2006). The utility of genetically informative data in the study of development. In S. Bergeman & S. M. Boker (Eds.), *Quantitative methodology in aging research* (pp. 269–327). Mahwah, NJ: Lawrence Erlbaum Associates.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling.* (VCU Box 900126, Richmond, VA 23298: Department of Psychiatry. 6th Edition)

Rakov, T., & Marcoulides, G. A. (2006). *A first course in structural equation modeling.* Mahwah, NJ: Laurence Erlbaum Associates.

Questionnaire

This nature of the topics presented in this course will, to some degree, be tailored to the interests and preparedness of the students. Please help me by filling in the following few questions.

1. Name:
2. Email Address:
3. Advisor:
4. Area (Social, Developmental, etc):
5. Year (1st, 2nd, etc):
6. Which undergraduate math, statistics, methodology classes have you taken?
7. Which graduate quantitative and methodology classes have you taken?
8. What projects are you currently working on?
9. Briefly, why are you taking graphical data analysis?

Please rate your interest in the following topics from on a scale from 1—5, where 0=No Interest, to 5=Extremely Interested. A sixth category labeled "?" is included in case you don't know enough about the topic to know whether you're interested. Two blank topics are included so that you can list other topics that might be of interest to you.

Topic	Interest (circle one)					
Cognitive Psychology	1	2	3	4	5	?
Clinical Psychology	1	2	3	4	5	?
Developmental Psychology	1	2	3	4	5	?
Quantitative Psychology	1	2	3	4	5	?
Social Psychology	1	2	3	4	5	?
Community Psychology	1	2	3	4	5	?
	1	2	3	4	5	?
	1	2	3	4	5	?