Demystifying Structural Equation Modeling

Jonathan Amburgey and Sean Raleigh, Westminster College

2022-05-06

Contents

Introduction

Welcome to our book on structural equation modeling!

Some history

In 2016, Jonathan and Sean embarked upon a bold experiment, asking the question, "Is is possible to teach structural equation modeling (SEM) to undergraduates with little statistical background?" To make things even more exciting, we attempted to do so in a special topics course lasting only one month during our May Term at Westminster College (Salt Lake City, UT).

In such an endeavor, we had to temper our expectations, of course. The goal was not to produce competent practitioners who would subsequently go on to do serious research using SEM techniques. We were quite happy that, at the end of May, we had undergraduates who were able to put together a simple final project that required them to find some data, posit a model, fit the model in R, interpret the output, and check a few model fit statistics. Some exposure to the topic and some appreciation of its power were satisfying enough. In fact, we think we got a little more out of it than that: we are reasonably confident that most of our students had developed—at that point right after taking the course—the ability to read a research article with an SEM model and have at least some idea what the article was talking about. We called it a win!

We repeated the experiment with some modifications to our materials and pedagogy in 2018. By that point, it was clear that finding textbooks and articles to assign to students was challenging. There are some great books out there, but they are mostly aimed at graduate students. Even the ones labeled "introductory" were often far from that for the typical undergraduate with limited statistical training.

We decided that we could write our own textbook that would fill this hole in the literature. The book that follows is the fruit of our efforts.

Sean was granted sabbatical in Spring 2020 and proposed to use that time to start writing the book in preparation for running the May Term course again in May 2020. And, well, we all know how that went…

Once the pandemic subsided enough for us to offer the course in person again, we attempted it again in May, 2022. [TO BE CONTINUED]

Our philosophy

As we mentioned before, our motivation for writing the book was driven by the difficulty we had finding readings for the students. Perhaps that begs the question, should one even try teaching a topic as "difficult" or "advanced" as structural equation modeling to the audience we had? To be sure, the traditional approaches already on the market seem to assume a lot more background that we had at our disposal. And the books that claim to assume less background…well, sometimes they require more than they let on.

The prerequisite for our class is an intro stats class that covers pretty standard material for such a course: hypothesis testing and confidence intervals for one and two proportions, one and two means (and paired means), ANOVA, chi-squared, and simple linear regression. We also benefit in that our intro course introduces students to R. (For those lacking R background, the first four modules here [FIX THIS LINK ONCE THE DATA 220 BOOK IS ALSO FULLY ONLINE] should suffice as a basic introduction to R and R Markdown sufficient for success in this course.)

To respect our students, we made some very deliberate choices about the way our book would be structured.

- *Make the book free and open source.*
  - Students have enough trouble in their lives and shouldn't be exposed to the extortionate practices of most textbook publishers. Not only is this book freely available online, it's also published under a permissive open source license that allows folks to "use, copy, modify, merge, publish, distribute, sublicense, and/or sell" their own versions of the book as desired. Furthermore, any copy of the book must also abide by the same open license. (Link to the MIT license.) So our book is both *libre* and *gratis* (or, in more common parlance, "free as in speech" and "free as in beer").
- *Start from scratch.*
  - Explain everything from the beginning in terms that are as simple as possible. Some of the first few chapters may look like review for students. Even if it is, of course, that review gives students confidence to tackle upcoming new material. But you might be surprised at some of the novel ways we explain seemingly familiar concepts. All explanations have an eye toward direct application in later chapters, so what might seem a little idiosyncratic at first actually has some motivation for smoothing the pathways into later concepts.

- *Incorporate active learning into everything.*

  – The chapters are structured to work as templates for classroom experiences. They intersperse conceptual explanation with activities designed to reinforce those concepts and lead students to important conclusions.

- *Do the math and do it well.*

  – One common thread we see in a lot of SEM books is a tendency to sweep most of the math under the rug. The intention comes from a good place; mathematics can appear intimidating and, therefore, may seem to serve as a deterrent to learning. To be sure, there are some complex mathematical ideas in SEM that are inaccessible to our audience. At the same time—and, in fairness, this may be due to Sean's bias as a mathematician—we truly believe that the mathematics, carefully explained, can illuminate student understanding. The more mathy sections may need additional instructor support for students without a strong math background. But all it takes is some relatively straightforward algebra to nail down some concepts that most books ignore. A good example of this is investing time in the rules for manipulating variances and covariances. This allows students to calculate the "model-implied matrix" that is only cryptically referenced in most textbooks. However, we do skip the math sometimes. For example, a lot of the math behind model fit indices is left unexplained. At the very least, we hope to be transparent about our choices to include or exclude certain mathematical details.

- *Use "easy" data.*

  – Finding data is hard, so we rely a lot on data sets that other textbooks and R package authors make available (with due attribution, of course). To keep things simple for this course, we work almost exclusively with numerical (quantitative) data. [MODIFY THIS IF WE END UP WORKING WITH BINARY CATEGORICAL EXOGENOUS VARIABLES (CODED 0/1) AT SOME POINT.]

- *Be careful about diagrams.*

  – Learning about complex models induces a sizable cognitive load. Shortcuts in diagrams tend to confuse students. For example, if error terms are truly latent variables, they should be drawn as circles or ellipses and not hidden, even if an advanced practitioner "knows" they're there. Variances and covariances among exogenous variables should always appear as well. We take the time to build up a consistent pictographic representation of every part of a model. (Each chapter is introduced with an archetypal diagram that illustrates that chapter's content.) Then we stick to that representation throughout the book.

- *Be careful about notation.*
    - While it may be the industry standard, LISREL notation is needlessly complex for undergraduate students. We take a consistent and simple approach to notation that represents all variables using UPPERCASE names and all path values using lowercase names. Abstract variables tend to called something like X when exogenous and Y when endogenous. When variables come from real data, they have contextually meaningful names. For those interested in reading the research literature, we have included an appendix describing LISREL notation.

Course structure

We use this book to teach a 2-credit-hour course. (Even though it's a special topics course in our May Term, the number of contact hours for students is equivalent to a semester-long 2-credit-hour course.)

[ADD INFO HERE AS WE DECIDE HOW MUCH IS REASONABLE TO COVER. IF WE WANT THE BOOK TO BE USABLE IN A 4-CREDIT-HOUR COURSE, WHAT ADDITIONAL MATERIAL SHOULD WE CONSIDER INCLUDING?]

Chapter 1

Variables and measurement

Chapter 2

Variance

Chapter 3

Covariance

Chapter 4

Simple regression

Chapter 5

Multiple regression

Chapter 6

Mediation

Chapter 7

Path analysis

Chapter 8

Latent variables

Chapter 9

Confirmatory factor analysis

Chapter 10

Structural equation models

Chapter 11

Structural causal models

Appendix (LISREL notation)