

Preface

What is assumed?

There are two ways of learning about structural equation modeling (SEM). The one I have chosen for this book is best described by an old advertising tag for a sport shoe company: “Just do it”. My approach could be called kinetic learning because it is based on the tactile experience of learning about SEM by using Stata to estimate and interpret models. This means you should have Stata open while you read this book; otherwise, this book might help you go to sleep if you try to read it without simultaneously working through it on your computer. By contrast, if you do work through the examples in the book by running the commands as you are reading, I hope you develop the same excitement that I have for SEM.

The alternative approach to learning SEM is to read books that are much more theoretical and may not even illustrate the mechanics of estimating models. These kinds of books are important, and reading them will enrich your understanding of SEM. This book is not meant to replace those books, but simply to get you started. My intent is for you to work your way through this book sequentially, but I recognize that some readers will want to skip around. I am hopeful that after you have been through this book once, you will want to return to specific chapters to reference techniques covered there. To facilitate this, each chapter includes some repetition of the most salient concepts covered in prior chapters. There is also a detailed index at the end of the book.

What background is assumed? A person who has never used Stata will need some help getting started. A big part of Stata’s brilliance is its simplicity, so a few minutes of help will get you up and ready for what you need to know about Stata. If you are new to Stata, have a friend who is familiar with the program show you the basics. If you have read my book *A Gentle Introduction to Stata* (2012a), you are ahead of the game. If you have any experience using Stata, then you are in great shape for this book. If you are a longtime Stata user, you will find that parts of this book explain things you already know.

To get the most out of this book, you need to have some background in statistics with experience in multiple regression. If you know path analysis, you will find the SEM approach to path analysis a big improvement over traditional approaches; however, the material on path analysis has been written for someone who has had very little exposure to path analysis. Even though the first chapter begins by covering how factor analysis has been used traditionally, a background in factor analysis is less important than having

had some exposure to multiple regression. The first chapter shows how confirmatory factor analysis adds capabilities to move beyond the traditional approach—you may never want to rely on alpha and principal component factor analysis again for developing a scale. I have covered enough about the traditional applications of factor analysis that you will be okay if you have had little or no prior exposure to factor analysis.

What will I learn?

We will explore many of the most widely used applications of SEM. We will begin with how to estimate a confirmatory factor analysis model—this is the measurement model part of SEM. This chapter includes parceling as a way to handle a large number of items for one or more of your factors. Next, we will cover path models using SEM—this is the structural model part of SEM. This chapter also introduces nonrecursive path models. We then put these two components together to introduce the full structural equation model. This chapter on the full model includes a number of specialized actions, such as equality constraints. With this foundation, we move on to a chapter on growth curves and conclude with a chapter on multiple-group analysis.

The book has two appendixes. Appendix A shows you how to use Stata’s graphical user interface (GUI) to draw and estimate models with Stata’s SEM Builder. It would be very useful to begin here so that you are familiar with the SEM Builder interface. If you have no background in SEM, you will not understand how to interpret the results you generate in appendix A, but this is not the point. Appendix A is just there to acquaint you with the SEM Builder that Stata introduced in version 12 and enhanced in version 13. How the interface works is the focus of appendix A. In the text, I use this GUI fairly often, but the focus is on understanding why we are estimating models the way we do and how we interpret and present the results. All the figures presented in this book were created using the SEM Builder, which produces publication-quality figures—far better than what you can draw with most other software packages that produce “near” publication-quality figures.

Appendix B shows you how to work with summary data (means, standard deviations, correlations) that are often reported in published works. You will be able to fit most models with these summary statistics even if you do not have the real data. This feature is great when you read an article and would like to explore how alternative models might be more appropriate. Many articles include a correlation matrix along with standard deviations and means. If these are not included, it is easier to request them from the author than it is to request the author’s actual data.

In addition to the two main appendixes, the first two chapters each have their own appendix that briefly describes using the SEM Builder for the models estimated in that chapter.

How do I use this book?

The chapters are intended to be read in the order they appear, and you should follow along with your reading on your computer. (It does not matter whether you use a Windows, Mac, or Unix operating system because the Stata commands and results will be identical.) If you have no background in SEM or factor analysis, take your time reading chapter 1. If you are comfortable with SEM and factor analysis, you should still go over chapter 1 enough to get a feel for how Stata works with the measurement model.

Though the chapters are fairly long, they are broken up into more manageable sections. If you are like me, once you know the commands I cover, you will have enough on your plate that you will forget the specifics before you need to fit a particular type of model. The sections in each chapter build on each other but are sufficiently independent that you should find them useful as a reference. Someday you will want to estimate a nonrecursive path model or a mediation model; you can easily find the section covering the appropriate model and come back to it. At the same time, this book does not attempt to compete with Stata's own *Structural Equation Modeling Reference Manual*, or [SEM]; I only cover a widely used subset of the options and postestimation commands available in Stata's SEM package.

What resources are available?

To facilitate the kinetic part of learning, you can download all the data used in this book as well as the Stata programs, called do-files, that fit every model. In the Command window, type the following:

```
. net from http://www.stata-press.com/data/dsemusr/  
. net describe dsemusr  
. net get dsemusr
```

When you run these three commands, you do not type the initial period and space, called the dot prompt. A convention in all Stata documentation and output in the Results window is to include the dot prompt as a prefix to each command, but you need only type the command itself.

There are several varieties of Stata software, and all of these are able to run the models described in this text. I focus on the Windows and Mac operating systems, and I show when there are slight differences in how they work in the GUI. The Unix GUI is very similar to the Windows GUI. The same Stata do-files run on all operating systems, though the systems differ slightly in how the file structure is organized.

One variety of Stata is called Small Stata. This is full featured and is small only in the sense of being limited in the number of observations (1,200) and variables (99) it can handle. Because a few of the datasets I use have more than 1,200 observations, I have made up smaller datasets that will work using Small Stata. You can obtain these datasets by entering the following in the Command window:

```
. net from http://www.stata-press.com/data/dsemusr/
. net describe dsemusr_small
. net get dsemusr_small
```

Using the Small Stata data, you will get somewhat different results for some models in the book simply because you will be using a smaller dataset. In addition, there are three models in chapter 4 that will not run using Small Stata.

At the end of each chapter, you will find some exercises that illustrate the material covered in the chapter. It is important to fit all the models in the text while you read the book because this reinforces what you are learning, as does typing in the commands yourself. The exercises extend this learning process by having you develop your own set of commands and models using the GUI system.

There is much more to SEM than could possibly be covered in a book this size. This book is intended to complement the material in the Stata manuals (over 11,000 pages of helpful information), which are available as PDF files when you install Stata. One way to access the [SEM] manual is to type `help sem` in the Command window of Stata. This opens a help file. At the top left of the help file, the title ([SEM] `sem` and `gsem`) is highlighted in blue. Clicking on this blue link will open up the PDF file of the [SEM] manual.

This book contains a fairly detailed index. Although I have explanatory section headings and these are a good place to start searching for how to do something, the index is naturally much more detailed. You may need to find how to place equality constraints on a multiple-group analysis or on a pair of reciprocal paths. These are covered in very different sections of the book, and the index tells you where to find them. The index was written to be useful after you have read this book and are using it as a reference to guide you while fitting your own models on your own data.

Conventions

Typewriter font. I use a typewriter font when something would be meaningful to Stata as input. This would be the case for something you type in the Command window or in a do-file. If a command is separated from the main text, as in

```
. sem (compliance <- educ income gender)
```

a dot prompt will precede the command. I also use a typewriter font for all Stata results, variable names, folders, and filenames.

Bold font. I use a bold font for menu items and for buttons you click within a menu. The bold font helps distinguish the button from the text; for example, you might be instructed to click the **Adjust Canvas Size** button.

Slant font. I use a slant font when referring to keys on our keyboard, such as the *Enter* key.

Italic font. I use an *italic font* when referring to text in a menu that you need to replace with something else, such as an actual variable name.

Capitalization. Stata is case sensitive. The command `sem (compliance <- educ income gender)` will produce a maximum likelihood multiple regression. If you replace `sem` with `Sem`, Stata will report that it has no command called `Sem`. I will use lowercase for all commands and all observed variables. When I refer to latent variables, I will capitalize the first letter of the latent variable. A simple confirmatory factor analysis would be `sem (Alienation -> anomia isolate depress report)`. Only the latent variable, `Alienation`, is capitalized. The arrow indicates that observed variables measure how a person responds on an anomia scale labeled `anomia`, an isolation scale labeled `isolate`, a depression scale labeled `depress`, and a reported score from an observer labeled `report`. All four of these observed variables depend on their level of `Alienation`, the latent variable.