Wiley StatsRef:
Statistics Reference Online

Structural Equation Modeling: Software

By *Edward E. Rigdon*

Keywords: *structural equation modeling, software,* LISREL , EQS , *Amos,* M*plus,* M*x*

Abstract: The first widely distributed special-purpose software for estimating structural equation models appeared more than twenty years ago. Today, researchers can choose from a variety of packages that differ markedly in terms of their intellectual heritage, interface, statistical sophistication, flexibility, integration with other statistical packages, and price. Despite numerous advances, there is still substantial room for improvement.

Introduction

The first widely distributed special-purpose software for estimating **structural equation models** (SEM) appeared more than twenty years ago. The earliest packages were written with the Fortran programming language and designed to run on mainframe computers, in an age when computing time was a scarce commodity and users punched instructions on stacks of cards. Features and terminology reminiscent of that earlier time still survive in the latest versions of some older SEM packages. While SEM software has become much more sophisticated, professional and user-friendly over the years, SEM software still is not what it could be. There is still substantial room for improvement.

Today, researchers can choose from a variety of packages that differ markedly in terms of their intellectual heritage, interface, statistical sophistication, flexibility, integration with other statistical packages, and price. Researchers may find that a few key criteria will substantially narrow their list of possible choices. Then again, packages regularly add capabilities, often mimicking their competitors, so any purchase or use decision should be based on the latest information.

This overview includes only packages that are primarily designed to estimate conventional structural equation models, with extensions. It excludes packages which are designed for SEM variants such as Tetrad and **Partial Least Squares** (*see* **Structural Equation Modeling: Nontraditional Alternatives**). It also excludes sophisticated modeling packages such as aML (http://www.applied-ml.com/) or GLLAMM (http://www.gllamm.org/), which are powerful and flexible tools but which lack many of the features, such as a broad array of fit indices and other fit diagnostics from the SEM literature, that users would expect in a SEM package.

Georgia State University, Atlanta, GA, USA

This article was originally published online in 2005 in Encyclopedia of Statistics in Behavioral Science, © John Wiley & Sons, Ltd and republished in Wiley StatsRef: Statistics Reference Online, 2014.

Choices

At least a dozen packages are available whose primary or major purpose is the estimation of structural equation models. Typically, each package began as a tool created by leading SEM researchers to facilitate their own analyses, only gradually being adapted to the needs of a larger body of customers. This origin as a tool for a SEM expert may partly explain why these packages seem to offer so little help to the average or novice user. This article starts with a brief overview of most of the known packages, in alphabetical order.

The Amos package, written by James Arbuckle and distributed by SPSS (http://www.spss.com/amos/) was unusual when it first appeared. It was perhaps the first SEM package to be designed for a graphical computing environment, like Microsoft Windows®. Taking advantage of the environment's capabilities, Amos allowed users to specify models by drawing them, and offered users a set of drawing tools for the purpose. (Amos also includes a command language called Amos Basic.) Amos was also an early leader in implementing advanced missing data techniques. While these capabilities have been copied, to a certain degree, by other leading programs, Amos retains a reputation for being easy to use, and the package has continued to add innovations in this area.

Proc Calis, written by Wolfgang Hartmann, is a procedure within the SAS package (http://www.sas.com/). In the early 1980s, Proc Calis was arguably the most sophisticated SEM package available. Its ability to specify constraints on model parameters as nonlinear functions of other parameters was instrumental in allowing researchers to model quadratic effects and multiplicative interactions between latent variables [2] (*see* **Structural Equation Modeling: Nonstandard Cases**). Over the intervening years, however, Proc Calis has not added features and extended capabilities to keep pace with developments.

EQS, written by Peter Bentler, has long been one of the leading SEM packages, thanks to the extensive contributions of its author, both to the program and to the field of SEM (http://www.mvsoft.com/). EQS was long distinguished by special features for dealing with nonnormal data, such as a kurtosis-adjusted χ^2 statistic [4], and superior procedures for modeling ordinal data (*see* **Ordinal Regression Models**).

LISREL, written by Karl Jöreskog and Dag Sörbom, pioneered the field of commercial SEM software, and it still may be the single most widely used and well-known package (http://www.ssicentral.com/). Writers have regularly blurred the distinction between LISREL as software and SEM as statistical method. Along with EQS, LISREL was an early leader in offering procedures for modeling ordinal data. Despite their program's advantage in terms of name recognition, however, the pressure of commercial and academic competition has forced the authors to continue updating their package. LISREL's long history is reflected both in its clear Fortran legacy and its enormous worldwide knowledge base.

Mplus (http://www.statmodel.com/), written by Linda and Bengt Muthén, is one of the newest entrants, but it inherits a legacy from Bengt Muthén's earlier program, LISCOMP. Besides maintaining LISCOMP's focus on nonnormal variables, Mplus has quickly built a reputation as one of the most statistically sophisticated SEM packages. Mplus includes tools for finite mixture modeling and **latent class analysis** that go well beyond conventional SEM, but which may point to the future of the discipline.

Mx (http://griffin.vcu.edu/mx/), written by Michael Neale, may be as technically sophisticated as any product on the market, and it has one distinguishing advantage: it's free. The software, the manual, and a graphical interface are all available free via the Internet. At its core, perhaps, Mx is really a matrix algebra program, but it includes everything that most users would expect in a SEM program, as well as leading edge capabilities in modeling incomplete data and in finite mixture modeling (*see* **Finite Mixture Distributions**).

SEPATH, written by James Steiger, is part of the Statistica statistical package (http://www.statsoftinc.com/products/advanced.html#structural). SEPATH incorporates many of Steiger's innovations relating to the analysis of correlation matrices. It is one of the few packages that automatically provides correct estimated standard errors for parameter estimates from SEM analysis of a Pearson correlation matrix (*see*

Correlation and Covariance Matrices). Most packages still produce biased estimated standard errors in this case, unless the user takes some additional steps.

Other SEM packages may be less widely distributed but they are still worth serious consideration. LINCS, written by Ronald Schoenberg, and MECOSA, written by Gerhard Arminger, use the GAUSS (http://www.aptech.com/) matrix algebra programming language and require a GAUSS installation. RAMONA, written by Michael Browne and Gerhard Mels, is distributed as part of the Systat statistical package (http://www.systat.com/products/Systat/productinfo/?sec=1006). SEM, by John Fox, is written in the open-source R statistical programming language (http://socserv.socsci.mcmaster.ca/jfox/Misc/sem/index.html). Like Mx, Fox's SEM is free, under an open-source license.

Finally, STREAMS (http://www.mwstreams.com/) is not a SEM package itself, but it may make SEM packages more user-friendly and easier to use. STREAMS generates language for, and formats output from, a variety of SEM packages. Users who need to take advantage of specific SEM package capabilities might use STREAMS to minimize the training burden.

Criteria

Choosing a SEM software package involves potential tradeoffs among a variety of criteria. The capabilities of different SEM packages change regularly – Amos, EQS, LISREL, Mplus, and Mx have all announced or released major upgrades in the last year or two–so it is important to obtain updated information before making a choice.

Users may be inclined to choose a SEM package that is associated with a more general statistical package which they already license. Thus, licensees of SPSS, SAS, Statistica, Systat, or GAUSS might favor Amos, Proc Calis, SEPath, RAMONA or MECOSA, respectively. Keep in mind that leading SEM packages will generally have the ability to import data in a variety of file formats, so users do not need to use the SEM software associated with their general statistical package in order to be able to share data across applications.

Many SEM packages are associated with specific contributors to SEM, as noted previously. Researchers who are familiar with a particular contributor's approach to SEM may prefer to use the associated software. Beyond general orientation, a given contributor's package may be the only one that includes some of the contributor's innovations. Over time, successful innovations do tend to be duplicated across programs, but users cannot assume that any given package includes tools for dealing with all modeling situations. Currently, for example, Amos does not include any procedures for modeling **ordinal data**, while Proc Calis does not support multiple group analysis, except when all groups have exactly the same sample size. Some features are largely exclusive, at least for the moment. Mplus and Mx are the only packages with tools for mixture modeling and latent class analysis. This allows these programs to model behavior that is intrinsically categorical, such as voting behavior, and also provides additional options for dealing with heterogeneity in a data set [3]. Similarly, only LISREL and Mplus have procedures for obtaining correct statistical results from modeling ordinal data without requiring exceptionally large sample sizes.

'Ease of use' is a major consideration for most users, but there are many different ways in which a package can be easy to use. Amos pioneered the graphical specification of models, using drawing tools to specify networks of observed and latent variables. Other packages followed suit, to one extent or another. Currently, LISREL allows users to modify a model with drawing tools but not to specify an initial model graphically. On the other hand, specifying a model with drawing tools can become tedious when the model involves many variables, and there is no universal agreement on how to graphically represent certain statistical features of a model. Menus, 'wizards' or other helps may actually better facilitate model specification in such cases. Several packages allow users to specify models through equation-like statements, which can provide a convenient basis for specifying relations among groups of variables.

'Ease of use' can mean ease of interpreting results from an analysis. Researchers working with multiple, competing models can easily become overwhelmed by the volume of output. In this regard, Amos and Mx have special features to facilitate comparisons among a set of competing models.

'Ease of use' can also mean that it is easy to find help when something goes wrong. Unfortunately, no SEM package does an excellent job of helping users in such a situation. To a great extent, users of any package find themselves relying on fellow users for support and advice. Thus, the popularity and history of a particular package can be important considerations. On that score, veteran packages like LISREL offer a broad population of users and a deep knowledge base, while less widely used packages like Proc Calis present special problems.

'Ease of use' could also relate to the ability to try out a product before committing to a major purchase. Many SEM packages offer 'student' or 'demo' versions of the software, usually offering full functionality but only for a limited number of variables. Some packages do not offer a demo version, which makes the purchase process risky and inconvenient. Obviously, free packages like Mx do not require demo versions.

Finally, users may be concerned about price. Developing a full-featured SEM package is no small task. Add to that the costs of supporting and promoting the package, factor in the small user base, relative to more general statistical packages, and it is not surprising that SEM packages tend to be expensive. That said, users should give special consideration to the Mx package, which is available free via the Internet. Mx offers a graphical interface and a high degree of flexibility, although specifying some model forms will involve some programming work. Still, templates are available for conducting many types of analysis with Mx.

Room for Improvement

SEM software has come a long way from the opaque, balky, idiosyncratic, mainframe-oriented packages of the early 1980s, but today's SEM packages still frustrate and inconvenience users and fail to facilitate SEM analysis as well as they could, given only the tools available today. SEM packages have made it easier to specify basic models, but specifying advanced models may require substantial programming, often giving rise to tedious errors, even though there is very little user discretion involved, once the general form of the advanced model is chosen. Users sometimes turn to bootstrap and Monte Carlo methods, as when sample size is too small for stable estimation, and several packages offer this capability. Yet, users find themselves 'jumping through hoops' to incorporate these results into their analyses, even though, once a few choices are made, there really is nothing left but tedium. There is much more to be done in designing SEM packages to maximize the efficiency of the researcher.

Most SEM packages do a poor job of helping users when something goes wrong. For example, when a user's structural equation model is not identified – meaning that the model's parameters cannot be uniquely estimated – SEM packages either simply fail or point the user to one particular parameter that is involved in the problem. Pointing to one parameter, however, may direct the user more to the symptoms and away from the fundamental problem in the model. Bekker, Merckens and Wansbeek [1] demonstrated a procedure, implemented via a Pascal program, that indicates all model parameters that are not identified, but this procedure has not been adopted by any major SEM package.

Several packages have taken steps in the area of visualization, but much more could be done. Confirmatory factor analysis measurement models (*See* **Factor Analysis: Confirmatory**) imply networks of constraints on the patterns of covariances among a set of observed variables. When such a model performs poorly, there are sets of covariances that do not conform to the implied constraints. Currently, SEM packages will point to particular elements of the empirical covariance matrix where the model does a poor job of reproducing the data, and they will also point to particular parameter constraints that contribute to lack of fit. But again, as with identification, this is not the same as showing the user just how the data contradict the network of constraints implied by the model.

Packages could probably improve the advice that they provide about how a researcher might improve a poorly fitting model. Spirtes, Scheines, and Glymour [5] have demonstrated algorithms for quickly finding structures that are consistent with data, subject to constraints. Alongside the diagnostics currently provided, SEM packages could offer more insight to researchers by incorporating algorithms from the stream of research associated with the Tetrad program.

SEM users often find themselves struggling in isolation to interpret results that are somewhat vague, even though there is a large body of researcher experience with SEM generally and with any given program in particular. One day, perhaps, programs will not only generate results but will also help researchers evaluate those results, drawing on this body of experience to give the individual research greater perspective. This type of innovation – giving meaning to numbers – is emerging from the field of **artificial intelligence**, and it will surely come to structural equation modeling, one day.

Related Articles

Software for Statistical Analyses

References

[1] Bekker, P. A., Merckens, A. & Wansbeek, T. J. (**1994**). *Identification, Equivalent Models, and Computer Algebra*, Academic Press, Boston.
[2] Kenny, D. A. & Judd, C. M. (**1984**). Estimating the nonlinear and interactive effects of latent variables, *Psychological Bulletin 96*, 201–210.
[3] Muthén, B. O. (**1989**). Latent variable modeling in heterogeneous populations: presidential address to the psychometric society, *Psychometrika 54*, 557–585.
[4] Satorra, A. & Bentler, P. M. (**1988**). Scaling corrections for chi-square statistics in covariance structure analysis, in 1988 Proceedings of the Business and Economic Statistics Section of the American Statistical *Association*, Washington, DC, 308–313.
[5] Spirtes, P., Scheines, R. & Glymour, C. (**1990**). Simulation studies of the reliability of computer-aided model specification using the TETRAD II, EQS, and LISREL Programs, *Sociological Methods & Research 19*, 3–66.