Structural Equation Modeling Using the sem Command and SEM Builder

Kristin MacDonald

Senior Statistician
StataCorp LP

2012 Stata Conference, San Diego

Outline

1. Terminology and model description
2. `sem` command syntax
3. SEM Builder
4. Tour of SEM models using the Builder

What is Structural Equation Modeling?

- SEM is class of statistical techniques that allows us to test hypotheses about relationships among variables.
- SEM encompasses other statistical methods such as correlation, linear regression, and factor analysis.
- SEM may also be referred to as Analysis of Covariance Structures. SEM fits models using the observed covariances and possibly means.

What is Structural Equation Modeling?

- SEM is class of statistical techniques that allows us to test hypotheses about relationships among variables.
- SEM encompasses other statistical methods such as correlation, linear regression, and factor analysis.
- SEM may also be referred to as Analysis of Covariance Structures. SEM fits models using the observed covariances and possibly means.

What is Structural Equation Modeling?

- SEM is class of statistical techniques that allows us to test hypotheses about relationships among variables.
- SEM encompasses other statistical methods such as correlation, linear regression, and factor analysis.
- SEM may also be referred to as Analysis of Covariance Structures. SEM fits models using the observed covariances and possibly means.

Types of variables

Exogenous vs. Endogenous

- Exogenous variables are not predicted by any other variables in the model.
- Endogenous variables are predicted by at least one other variable in the model.

Observed vs. Latent

- Observed variables are variables for which we have data (either observations in our dataset or matrices of covariances, means, etc.).
- Latent variables are unobserved variables and may represent hypothetical constructs, the true values of variables measured with error, unobserved heterogeneity, errors, and more.

Types of variables

Exogenous vs. Endogenous

- Exogenous variables are not predicted by any other variables in the model.
- Endogenous variables are predicted by at least one other variable in the model.

Observed vs. Latent

- Observed variables are variables for which we have data (either observations in our dataset or matrices of covariances, means, etc.).
- Latent variables are unobserved variables and may represent hypothetical constructs, the true values of variables measured with error, unobserved heterogeneity, errors, and more.

Models in the SEM framework

- linear regression
- ANOVA
- multivariate regression
- simultaneous equation models
- path analysis
- mediation analysis
- confirmatory factor analysis (CFA)
- higher order CFA models
- measurement models
- reliability estimation
- full structural equation models
- multiple indicators and multiple causes (MIMIC)
- latent growth curve models
- multiple group models

Models in the SEM framework

- linear regression
- ANOVA
- multivariate regression
- simultaneous equation models
- path analysis
- mediation analysis
- confirmatory factor analysis (CFA)
- higher order CFA models
- measurement models
- reliability estimation
- full structural equation models
- multiple indicators and multiple causes (MIMIC)
- latent growth curve models
- multiple group models

Models in the SEM framework

- linear regression
- ANOVA
- multivariate regression
- simultaneous equation models
- path analysis
- mediation analysis
- confirmatory factor analysis (CFA)
- higher order CFA models
- measurement models
- reliability estimation
- full structural equation models
- multiple indicators and multiple causes (MIMIC)
- latent growth curve models
- multiple group models

Models in the SEM framework

- linear regression
- ANOVA
- multivariate regression
- simultaneous equation models
- path analysis
- mediation analysis
- confirmatory factor analysis (CFA)
- higher order CFA models
- measurement models
- reliability estimation
- full structural equation models
- multiple indicators and multiple causes (MIMIC)
- latent growth curve models
- multiple group models

Mathematical notation for the model

$$Y = BY + \Gamma X + \alpha + \varsigma$$

where

Y is a vector of endogenous variables, both latent and observed

X is a vector of exogenous variables, both latent and observed

B and Γ are matrices of coefficients

α is a vector of intercepts

ς is a vector of error terms

Mathematical notation for the model

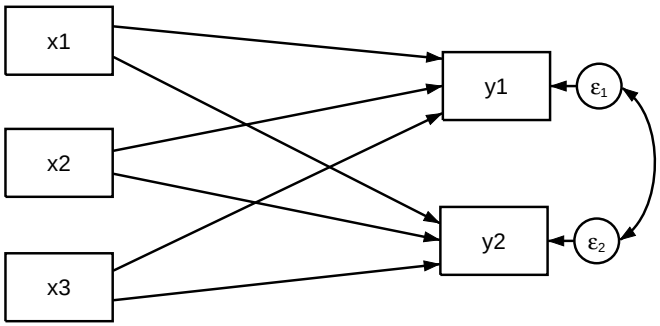
Also estimated are the variances of the exogenous variables and errors

$$\Phi = Var(X)$$
$$\Psi = Var(\varsigma)$$

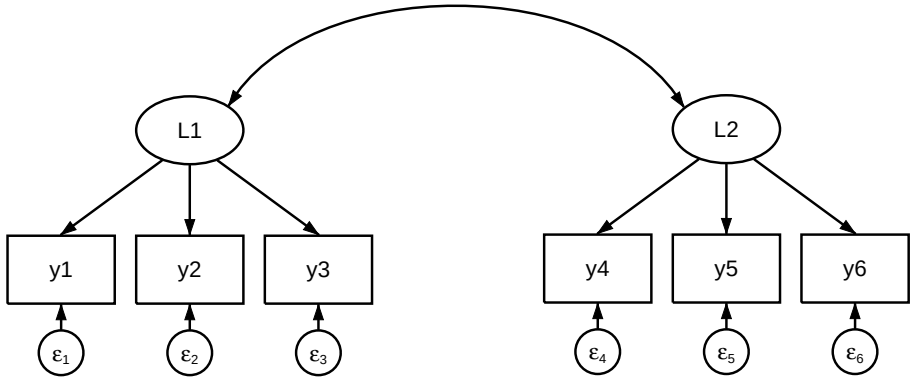
Path diagrams

- Observed variables represented by rectangles
- Latent variables represented by ovals
- Paths represented by arrows
- Covariances represented by curved lines with arrows at each end

Multivariate regression



Confirmatory factor analysis

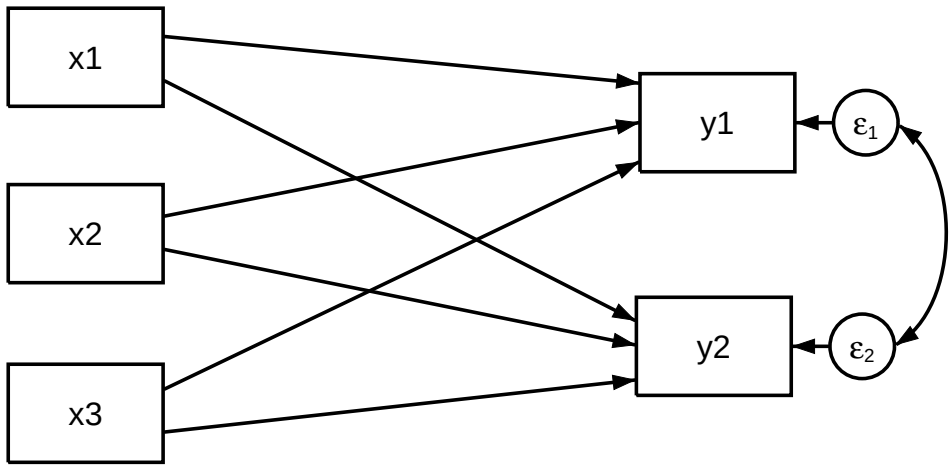


Syntax examples

```
. sem (x1 x2 x3 -> y1 y2), covstructure(e._En, unstructured)

. sem (y1 <- x1 x2 x3) (y2 <- x1 x2 x3), cov(e.y1*e.y2)

. sem (y1 <- x1) (y1 <- x2) (y1 <- x3)
       (y2 <- x1) (y2 <- x2) (y2 <- x3), cov(e.y1*e.y2)
```



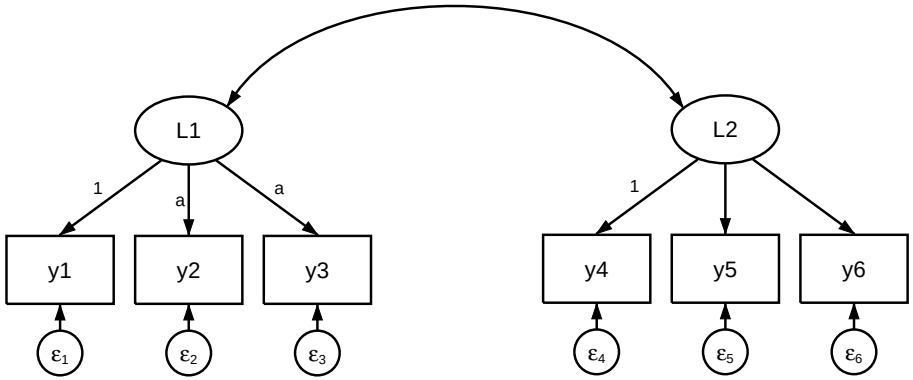
Syntax examples

- `sem` syntax mimics path diagrams, using arrows to specify presumed relationships among variables.
- Covariances can be specified using the `covariance()` or `covstructure()` option.
- Errors are referred to using the `e.` prefix with the endogenous variable's name.
- Arrows are allowed to face either direction. Paths can be specified individually or many paths can be specified at once.

Syntax examples

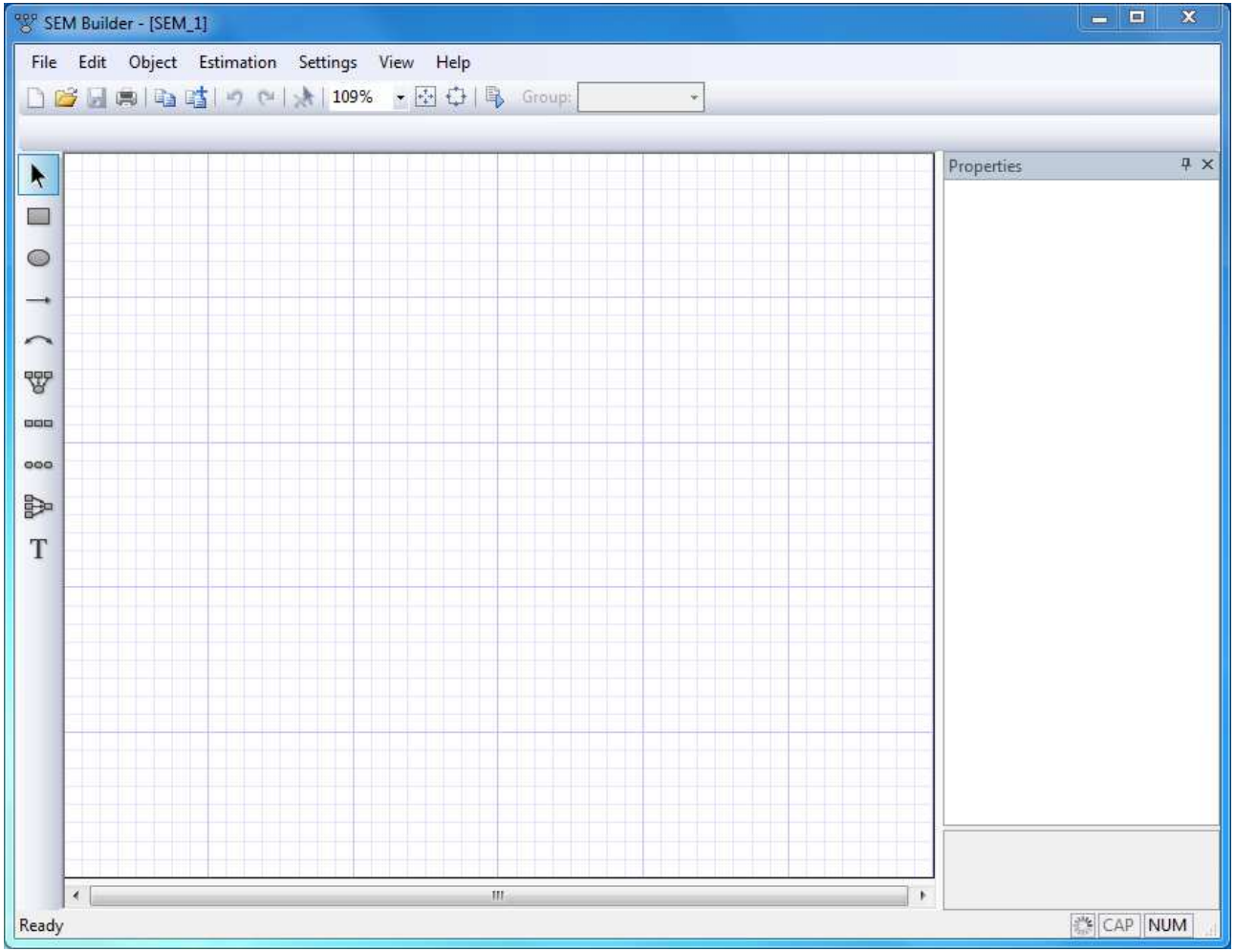
```
. sem (L1 -> y1@1 y2@a y3@a) (L2 -> y4@1 y5 y6), cov(L1*L2)

. sem (L1 -> y1 y2@a y3@a) (L2 -> y4 y5 y6)
```



- By default, sem assumes variables beginning with capital letters are latent variables.
- Constraints can be specified using @.
- All exogenous variables are assumed to covary when using the command syntax.

GUI for sem


SEM Builder - [SEM_1]
File Edit Object Estimation Settings View Help
109%
Group:
Properties
Ready
CAP
NUM

CFA example

- Using data from Holzinger and Swineford (1939), we fit a two factor model. In this example, we have latent variables representing verbal and spatial abilities. Each are measured using results from three tests.
- The -sem- command for this model

```
. sem (Verbal -> paragraph sentence wordc)
      (Spatial -> visual cubes paper)
```

Path analysis

- Path analysis is typically used to refer to models involving only observed variables.
- In an example from Fogarty, et al. (1999), the authors examined relationships between job satisfaction, occupational strain, occupational stress, coping abilities, negative affectivity, and positive affectivity.
- We could fit one of the proposed models using the following `sem` command.

```
. sem (stress <- na pa)
      (coping <- na pa stress)
      (strain <- na pa stress coping)
      (satisfaction <- na pa stress coping strain)
```

Full structural equation model

- A full structural equation model combines aspects of path analysis and confirmatory factor analysis. Latent variables are measured by observed variables and structural paths exist among variables.
- An example using data from Wheaton, et al. (1977) includes three latent variables with structural paths between latent variables and covariances among certain errors.
- The sem command for this model is

```
. sem (anomia67 pwless67 <- Alien67)
      (anomia71 pwless71 <- Alien71)
      (Alien67 <- SES)
      (Alien71 <- Alien67 SES)
      (SES -> educ occstat66),
              cov(e.anomia67*e.anomia71)
              cov(e.pwless67*e.pwless71)
```

Tips for production quality diagrams

- Use tools for adding sets of variables, measurement components, or regression components instead of adding individual variables to the path diagram.
- To align components that have already been added to the diagram, select **Object** → **Align** from the menu.
- After all variables are aligned properly, you may wish to allow the SEM Builder to automatically determine where arrows should connect to ovals and rectangles. Select **Settings** → **Automation** → **Attach based on position of variables**.

Tips for production quality diagrams

- Use tools for adding sets of variables, measurement components, or regression components instead of adding individual variables to the path diagram.
- To align components that have already been added to the diagram, select **Object** → **Align** from the menu.
- After all variables are aligned properly, you may wish to allow the SEM Builder to automatically determine where arrows should connect to ovals and rectangles. Select **Settings** → **Automation** → **Attach based on position of variables**.

Tips for production quality diagrams

- Use tools for adding sets of variables, measurement components, or regression components instead of adding individual variables to the path diagram.
- To align components that have already been added to the diagram, select **Object** → **Align** from the menu.
- After all variables are aligned properly, you may wish to allow the SEM Builder to automatically determine where arrows should connect to ovals and rectangles. Select **Settings** → **Automation** → **Attach based on position of variables**.

Tips for production quality diagrams

- Use the **Settings** → **Variables** and **Settings** → **Connections** menu options to make changes globally for the appearance of all variables or paths. It is usually easiest to start here and then make changes related to individual variables or paths if you still need to.

- Save your .stsem file so that it is easy to make modifications later. Also save your path diagram in other forms such as PDF, EPS, PNG to easily include in publications.

Tips for production quality diagrams

- Use the **Settings** → **Variables** and **Settings** → **Connections** menu options to make changes globally for the appearance of all variables or paths. It is usually easiest to start here and then make changes related to individual variables or paths if you still need to.

- Save your .stsem file so that it is easy to make modifications later. Also save your path diagram in other forms such as PDF, EPS, PNG to easily include in publications.

Using tags in the SEM Builder

- Tags can be included in labels and results to customize the path diagram.
- To see a list of available tags, type

```
. help sg_tags
```

- In the dialog boxes that are opened through **Settings** → **Variables** and **Settings** → **Connections**, the Label and Results tabs show some of these tags in use. If you make changes in the way labels or results are displayed using the drop down menus, the corresponding changes using tags will appear in the box next to the menu.

We have fit a few classic SEM models using the SEM Builder, but we have really just scratched the surface the capabilities of `sem`. In addition, we can

- obtain robust or cluster-robust standard errors using the `vce(robust)` and `vce(cluster)` options
- obtain bootstrap or jackknife standard errors
- use the `svy` prefix to take into account complex survey design
- instead of listwise deletion, use the `method(mlmv)` option to perform estimation using maximum likelihood estimation with missing at random data
- perform estimation using asymptotic distribution free (ADF) estimation rather than maximum likelihood estimation
- specify a known reliability of an observed variable using the `reliability()` option

Bollen, K. A. 1989. *Structural Equations with Latent Variables*. New York: Wiley.

Fogarty, G., A. M. Machin, M. Albion, L. Sutherland, G. L. Lalor, and S. Revitt. 1999. Predicting occupational strain and job satisfaction: The role of stress, coping, and positive and negative affectivity. *Journal of Vocational Behavior* 54:429—452.

Holzinger, K. J. and F. Swineford. 1939. A study in factor analysis: The stability of a bi-factor solution. *Supplementary Education Monographs, 48*. Chicago, IL: University of Chicago.

Kline, R. B. 2011. *Principles and Practice of Structural Equation Modeling*. 3rd ed. New York: Guilford Press.

Wheaton, B., B. Muthén, D. F. Alwin, and G. F. Summers. 1977. Assessing reliability and stability in panel models. In *Sociological Methodology 1977*, ed. D. R. Heise, 84–136. San Francisco: Jossey-Bass.