Manifest Variable Modeling with OpenMx

Steven M. Boker

Department of Psychology
University of Virginia

Structural Equation Modeling
Psyc–8501–001

Overview

- What is OpenMx?
  - Overview of the objects in OpenMx.
- Univariate regression model.
  - Standardized data using a covariance matrix.
  - Unstandardized data using Full Information Maximum Likelihood.
- Bivariate regression model.
  - Standardized data using a covariance matrix.
  - Unstandardized data using FIML.
- Four variable multiple regression.
  - Standardized data using a covariance matrix.
  - Unstandardized data using FIML.
- Multivariate regression.
  - Unstandardized data using FIML.

What is OpenMx?

- OpenMx is
  1. A free, full–featured, open source SEM package.
  2. Runs on Windows, Mac OS-X, and Linux.
  3. Runs inside the R statistical programming environment.
  4. Funded by the NIH Roadmap Initiative.
- OpenMx features:
  1. A new approach to model specification.
  2. Allows both path-style and matrix-style scripting.
  3. Flexible optimization including nonlinear constraints.
  4. Web–based forums, tutorials, and a wiki.
  5. Support for most popular types of modeling.
  6. Advanced features not found in other SEM packages.
  7. An *active* development team.

http://openmx.psyc.virginia.edu

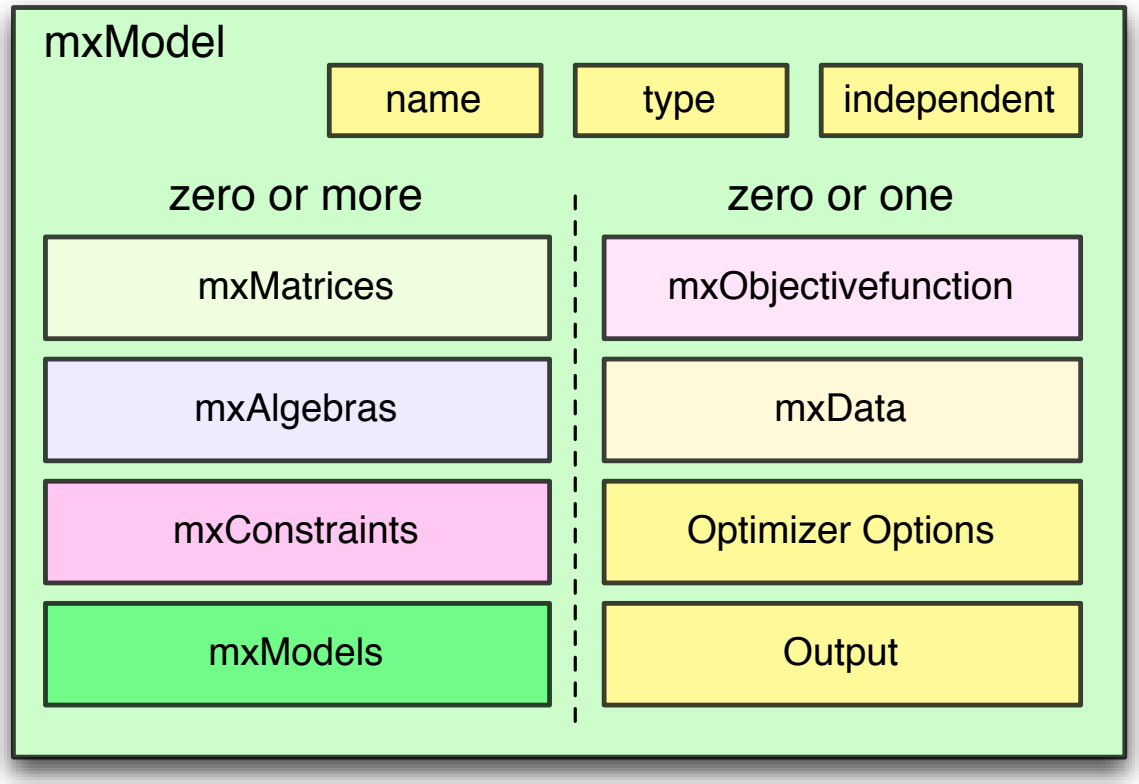
Why Open Source?

- Open source refers to a community–based approach to development of software.
- OpenMx is not a black box.
  - You can look at our code to see exactly how we calculate everything.
- OpenMx is built around the scientific model.
  - Acknowledgement of each other's work.
  - Contribution of one's own work to the benefit of all.
- We hope that OpenMx will provide quantitative graduate students a boost towards implementing their own ideas.
- You can use our code in your own projects!
  - Apache 2.0 License.

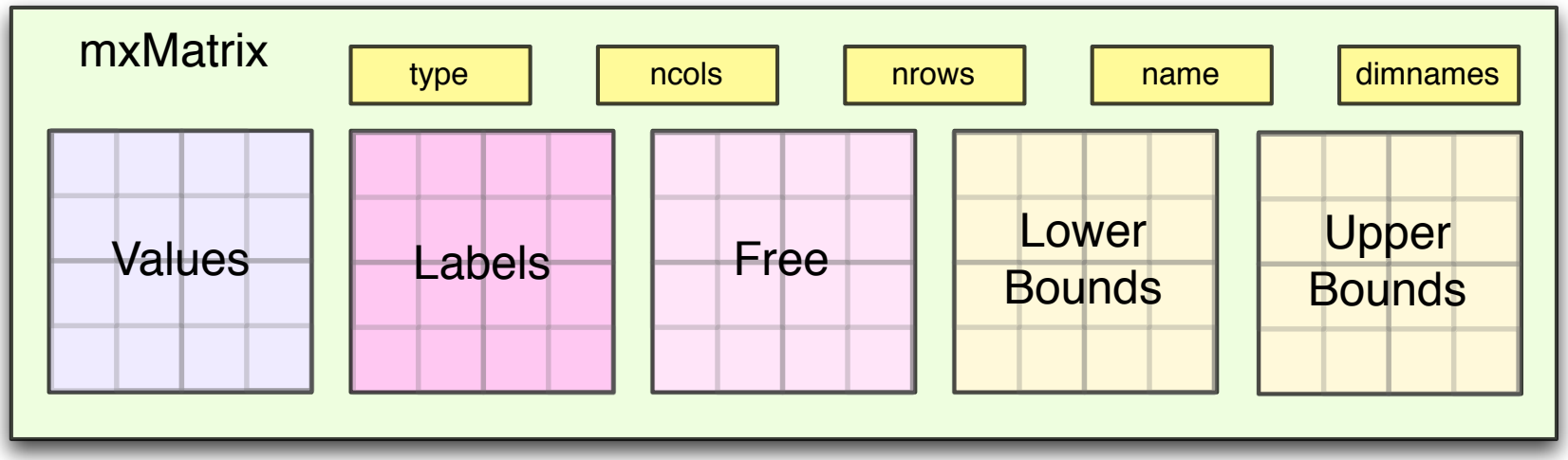
What Models Are Addressed?

- In the current beta version.
  - Multivariate Normal Structural Equation Models.
  - Multigroup Models, e.g. Behavior Genetic.
  - Full Information Maximum Likelihood.
  - Mixed Effects and Multilevel.
  - Multivariate Categorical Data with Thresholds.
  - Dynamical Systems Models.
  - Nonlinear Constraints.
  - Mixture Distribution Models.
  - Parallelizing Estimation.
  - User–supplied Matrix Algebra and Objective Functions.
  - Much, much more.
- Under development and coming soon.
  - Cross–Classified Multivariate Multilevel.
  - Multi–Chain Monte Carlo / Bayesian Estimation.

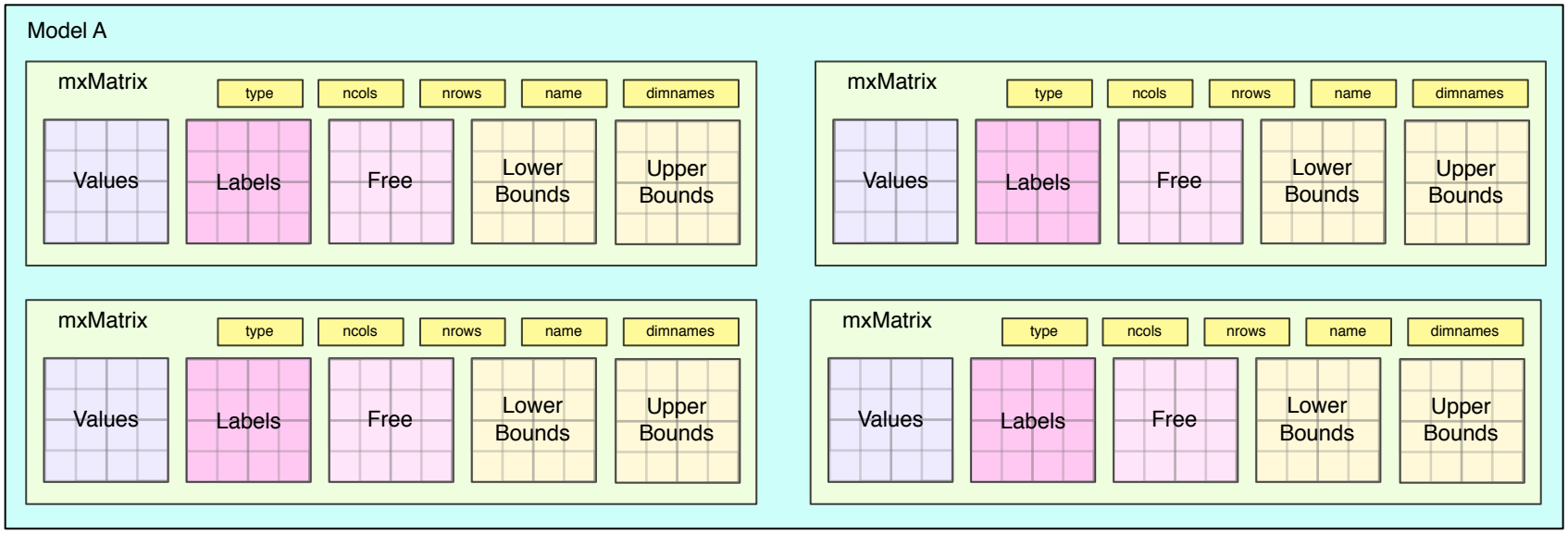
An MxModel Contains Objects and Other MxModels



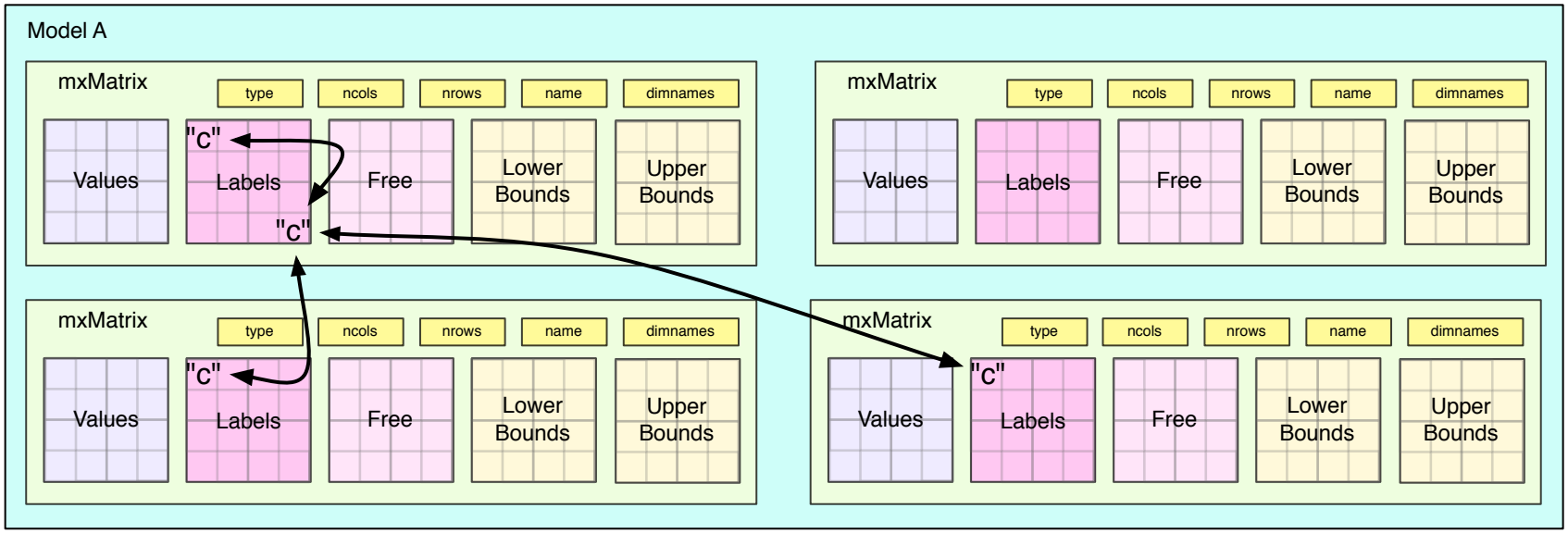
An MxMatrix Contains Values and Metainformation



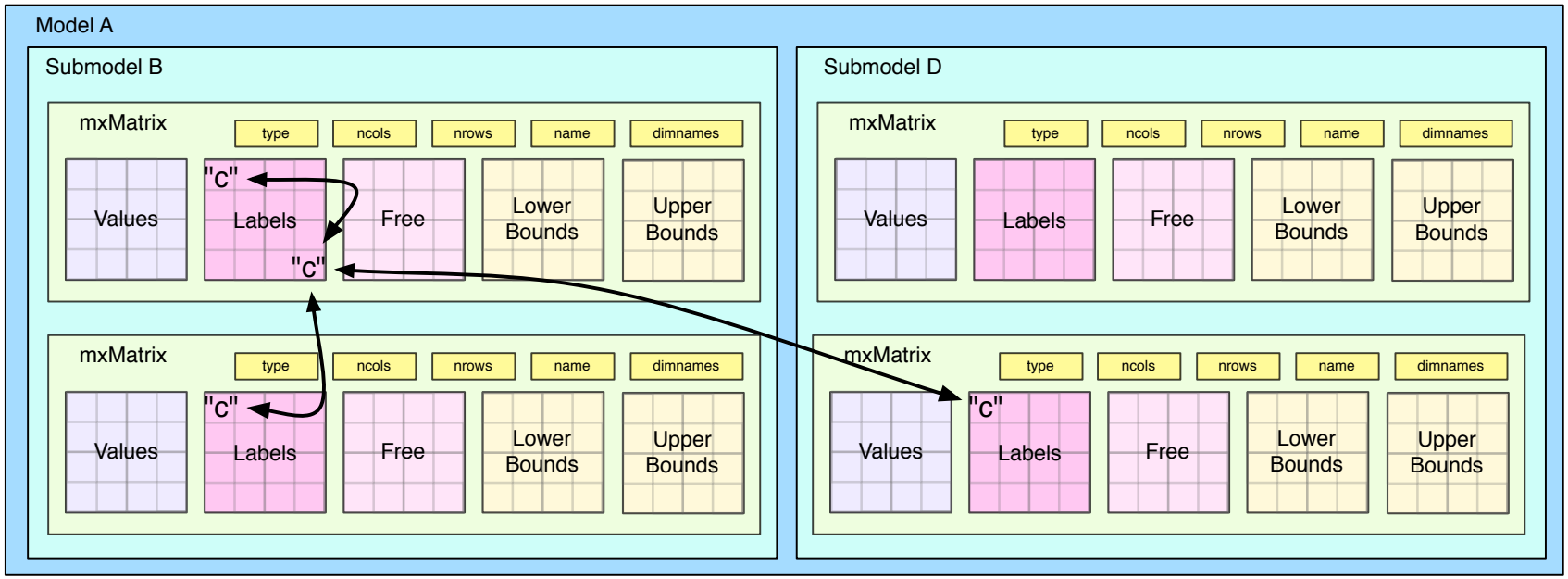
Many MxMatrices Can Be in an MxModel



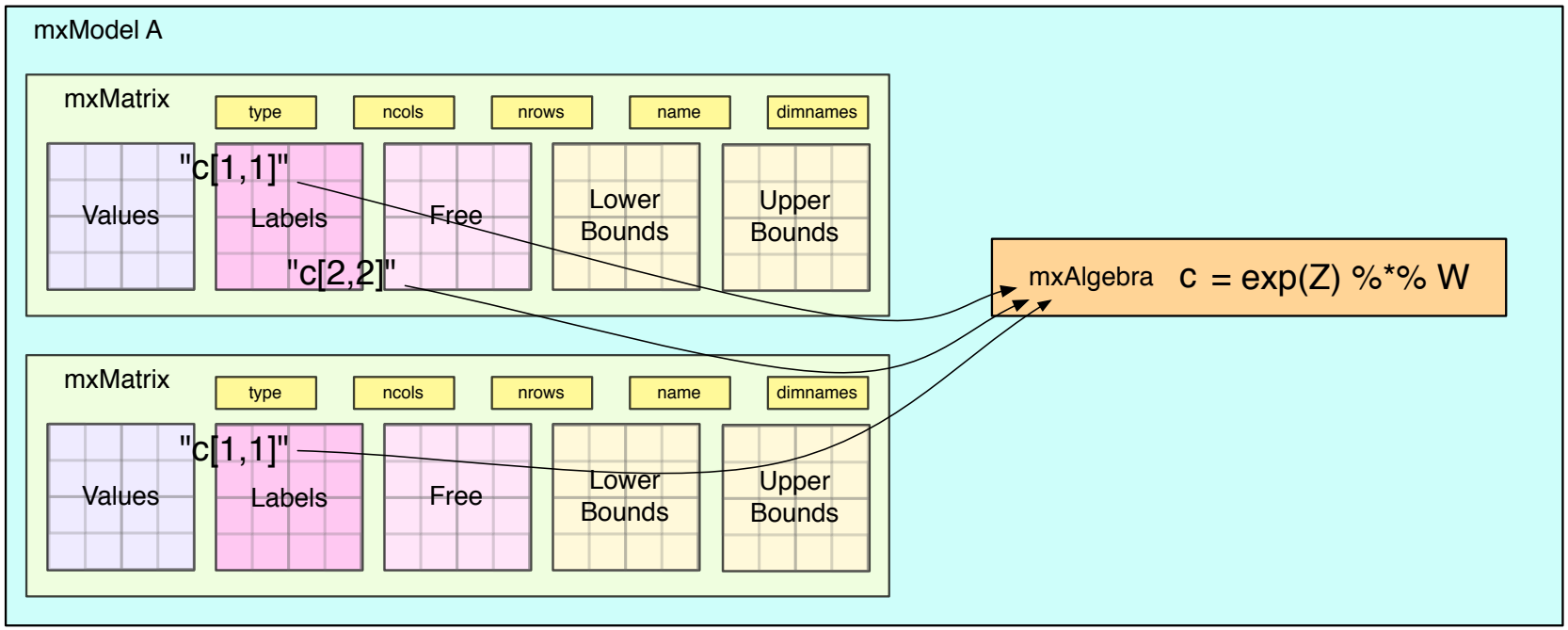
Labels Can Be Used For Equality Constraints



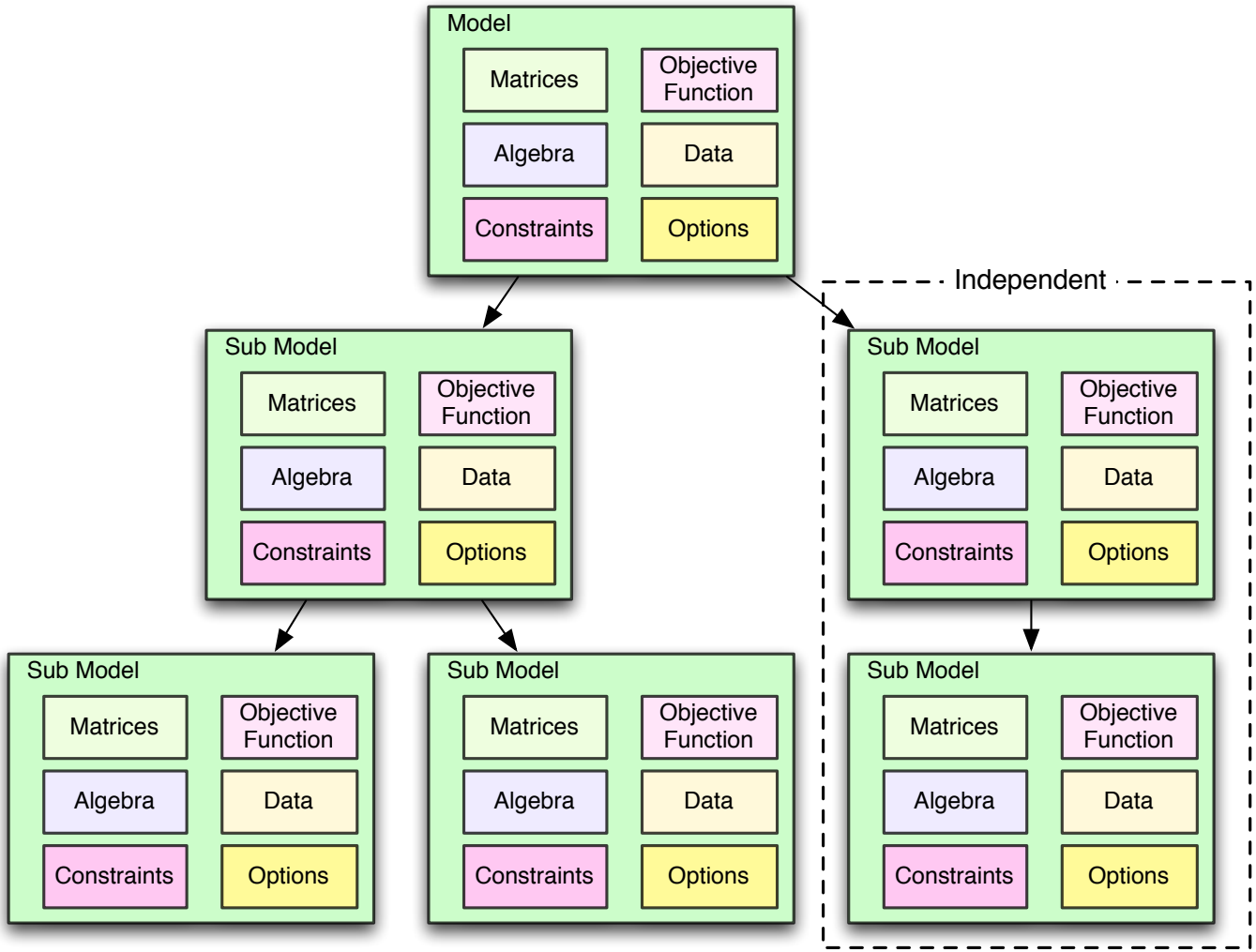
Equality Constraints can be Between Submodels



Labels Can Constrain to Algebraic Results



Models Can Be Hierarchically Structured



http://openmx.psyc.virginia.edu

- OpenMx is
  - Free.
  - Open Source (Apache 2.0 License).
  - Available now as a public beta test from the OpenMx website.
- OpenSEM is
  - A community for SEM modelers, teachers, and students.
  - A set of topic–based discussion forums.
  - Open to users of any software, not just OpenMx.
  - Free, with registration and login at the OpenMx website.
- The OpenMx team hopes you find our work useful.

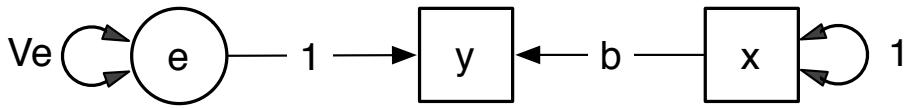
A Brief Introduction to R

- For those who have not used R before, I will give a very brief introduction to R by running quickly through an R script.
- The script is TW-IntroToR-20100210.R

R Simulation

- For our first examples, we will use some simulated regression data.
- If you want to peek at the last page of the novel, MultipleRegressionSim.R has the simulation.

Standardized Univariate Regression



$$y = bx + e$$

$$\mathbf{R} = \begin{bmatrix} 1.0 & b & 0 \\ b & b^2 + V_e & V_e \\ 0 & V_e & V_e \end{bmatrix}$$

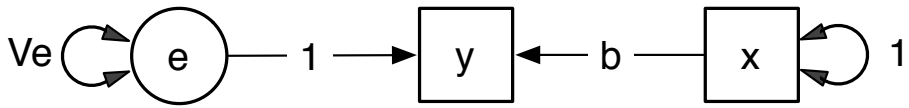
Running OpenMx

- The first example is UnivariateStd-OpenMx100214.R.

Evaluating the Univariate Result

- Is $x1$ related to y?
- What are the fit statistics and degrees of freedom?
- Let's try adding a latent variable for the error.
  - Does that change the answer at all?
- Let's look at the model matrices that OpenMx created.
- Let's look more closely at the summary().

Standardized Univariate Regression



$$y = bx + e$$

$$\mathbf{R} = \begin{bmatrix} 1.0 & b & 0 \\ b & b^2 + V_e & V_e \\ 0 & V_e & V_e \end{bmatrix}$$

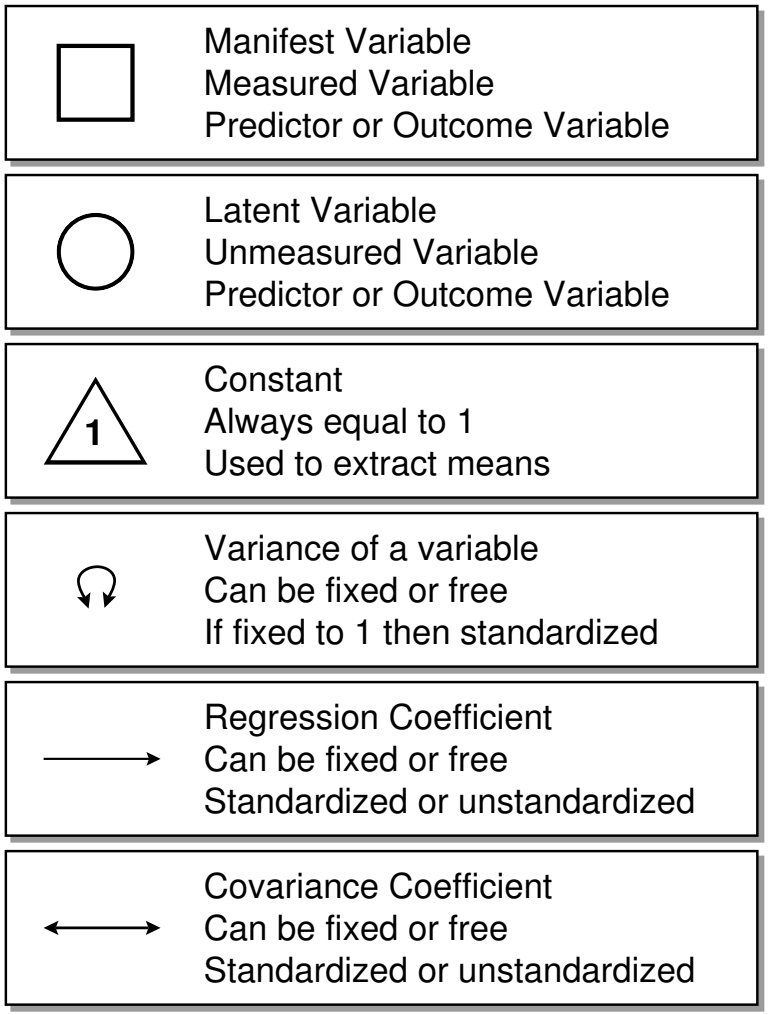
Parts of a Path–Style OpenMx Script

- Read in the data.
- Define the variables you want to use.
- Define the model.
  - Specify the manifest and latent variables.
  - Specify the regression paths.
  - Specify the variance paths.
  - Specify the covariance paths.
  - Specify the means paths.
  - Specify what data to use.
- Run the model.
- Look at the output.

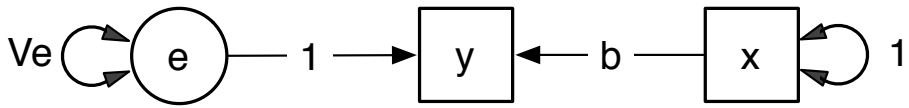
Full Information Maximum Likelihood

- We can represent a structural model as a *path diagram*.
- Squares are manifest variables.
- Circles are latent variables.
- **Triangles are constants.**
- Single headed arrows are regression coefficients.
- Double headed arrows are variances and covariances.

Specifying Means



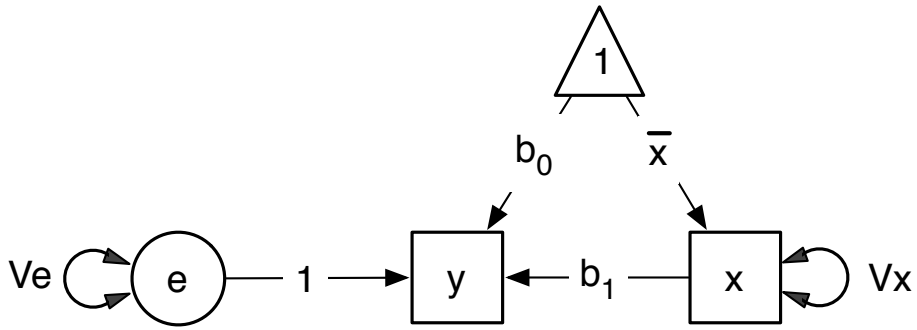
Standardized Univariate Regression



$$y = bx + e$$

$$\mathbf{R} = \begin{bmatrix} 1.0 & b & 0 \\ b & b^2 + V_e & V_e \\ 0 & V_e & V_e \end{bmatrix}$$

Full Information Maximum Likelihood



$$y = b_0 + b_1 x + e$$

$$\mathbf{R} = \begin{bmatrix} V_x & b_1 V_x & 0 \\ b_1 V_x & b_1^2 V_x + V_e & V_e \\ 0 & V_e & V_e \end{bmatrix}$$

$$\mathbf{M} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ 0 \end{bmatrix}$$

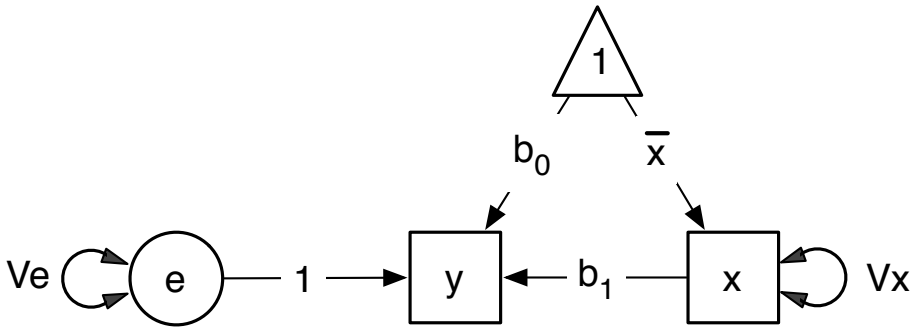
Running OpenMx

- The first FIML example is UnivariateRaw-OpenMx100214.R.

Evaluating the Univariate FIML Result

- Is $x1$ related to y?
- What are the fit statistics and degrees of freedom?
- Let's look at the model matrices that OpenMx created.
- Let's look more closely at the summary().
  - What is the mean of y?
  - What is the deal with those degrees of freedom?

Unstandardized Univariate Regression

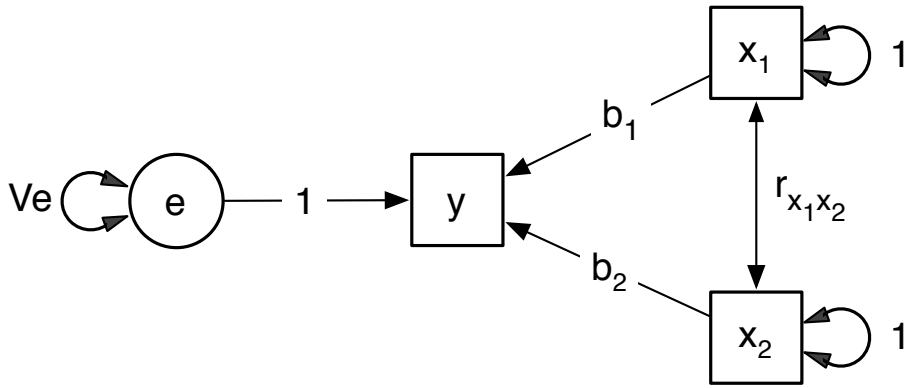


$$y = b_0 + b_1 x + e$$

$$\mathbf{R} = \begin{bmatrix} V_x & b_1 V_x & 0 \\ b_1 V_x & b_1^2 V_x + V_e & V_e \\ 0 & V_e & V_e \end{bmatrix}$$

$$\mathbf{M} = \begin{bmatrix} \bar{x} \\ \bar{y} \\ 0 \end{bmatrix}$$

Standardized Bivariate Regression



$$y_i = b_1 x_{i1} + b_2 x_{i2} + e_i$$

$$\mathbf{R} = \begin{bmatrix} 1 & r_{x_1x_2} & b_1 + b_2 r_{x_1x_2} & 0 \\ r_{x_1x_2} & 1 & b_2 + b_1 r_{x_1x_2} & 0 \\ b_1 + b_2 r_{x_1x_2} & b_2 + b_1 r_{x_1x_2} & b_1^2 + b_2^2 + 2 b_1 b_2 r_{x_1x_2} + V_e & V_e \\ 0 & 0 & V_e & V_e \end{bmatrix}$$

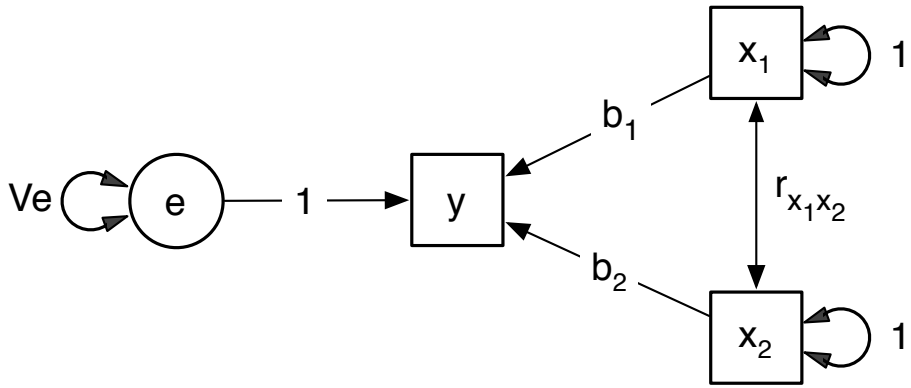
Running OpenMx

- The bivariate example is BivariateStd-OpenMx100214.R.

Evaluating the Bivariate Result

- Are $x1$ and $x2$ related to y?
- What are the fit statistics and degrees of freedom?
- Let's look at the model matrices that OpenMx created.
- Let's look more closely at the summary().

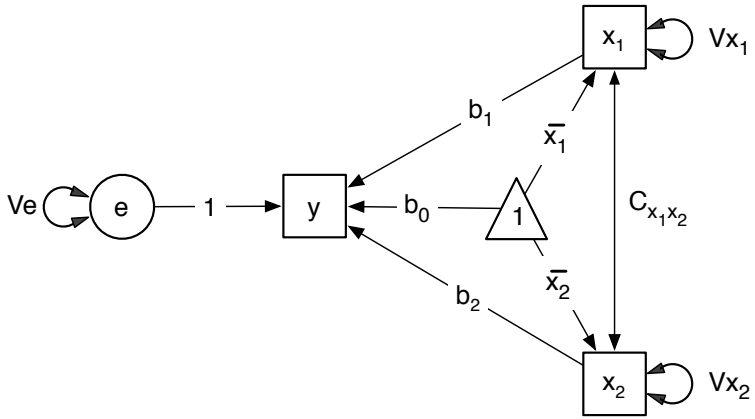
Standardized Bivariate Regression



$$y_i = b_1 x_{i1} + b_2 x_{i2} + e_i$$

$$\mathbf{R} = \begin{bmatrix} 1 & r_{x_1x_2} & b_1 + b_2 r_{x_1x_2} & 0 \\ r_{x_1x_2} & 1 & b_2 + b_1 r_{x_1x_2} & 0 \\ b_1 + b_2 r_{x_1x_2} & b_2 + b_1 r_{x_1x_2} & b_1^2 + b_2^2 + 2 b_1 b_2 r_{x_1x_2} + V_e & V_e \\ 0 & 0 & V_e & V_e \end{bmatrix}$$

FIML Bivariate Regression



$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i$$

$$\mathbf{R} = \begin{bmatrix} V_{x1} & C_{x_1 x_2} & b_1 V_{x1} + b_2 C_{x_1 x_2} & 0 \\ C_{x_1 x_2} & V_{x2} & b_2 V_{x2} + b_1 C_{x_1 x_2} & 0 \\ b_1 V_{x1} + b_2 C_{x_1 x_2} & b_2 V_{x2} + b_1 C_{x_1 x_2} & b_1^2 V_{x1} + b_2^2 V_{x2} + 2 b_1 b_2 C_{x_1 x_2} + V_e & V_e \\ 0 & 0 & V_e & V_e \end{bmatrix}$$

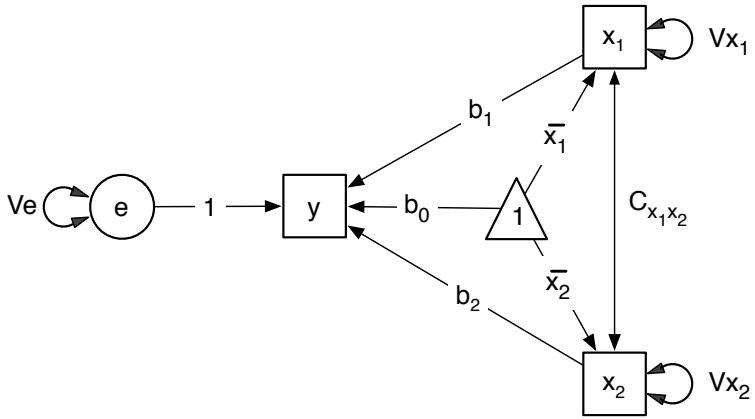
Running OpenMx

- The bivariate FIML example is BivariateRaw-OpenMx100214.R.

Evaluating the Bivariate FIML Result

- Are $x1$ and $x2$ related to y?
- What are the fit statistics and degrees of freedom?
- Let's look at the model matrices that OpenMx created.
- Let's look more closely at the summary().
  - Now, what is the mean of y?

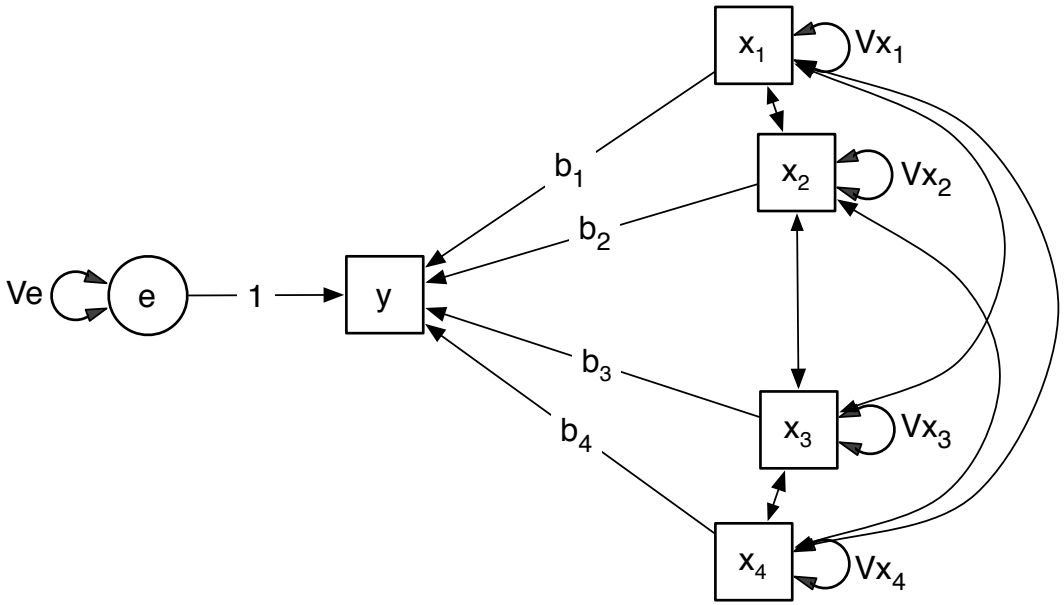
FIML Bivariate Regression



$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + e_i$$

$$\mathbf{R} = \begin{bmatrix} V_{x1} & C_{x_1 x_2} & b_1 V_{x1} + b_2 C_{x_1 x_2} & 0 \\ C_{x_1 x_2} & V_{x2} & b_2 V_{x2} + b_1 C_{x_1 x_2} & 0 \\ b_1 V_{x1} + b_2 C_{x_1 x_2} & b_2 V_{x2} + b_1 C_{x_1 x_2} & b_1^2 V_{x1} + b_2^2 V_{x2} + 2 b_1 b_2 C_{x_1 x_2} + V_e & V_e \\ 0 & 0 & V_e & V_e \end{bmatrix}$$

Standardized Multiple Regression



$$y_i = b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + e_i$$

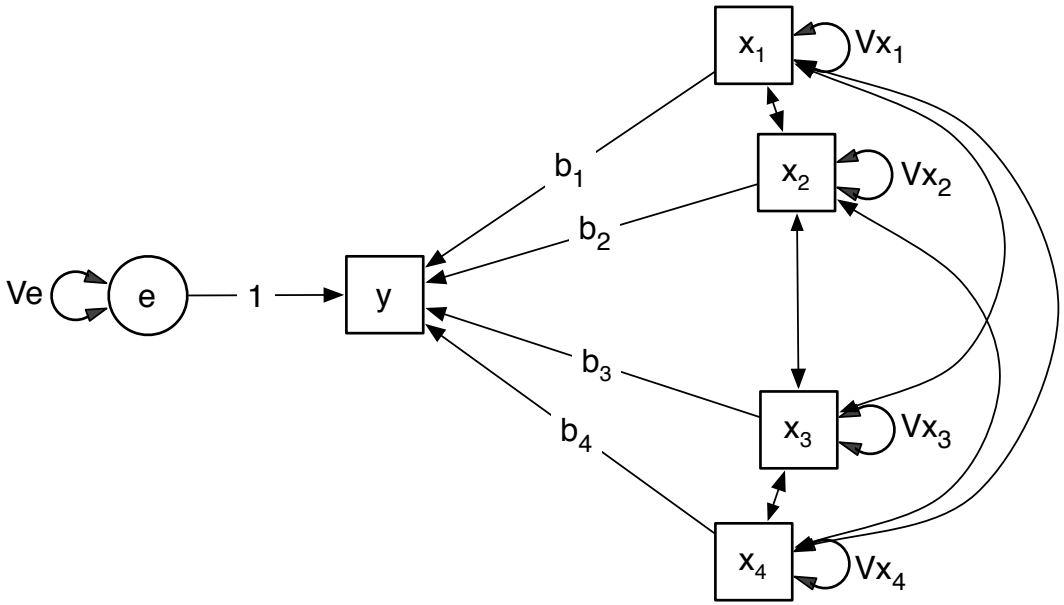
Running OpenMx

- The multiple regression standardized example is MultiRegStd-OpenMx100214.R.

Evaluating the Multiple Regression Result

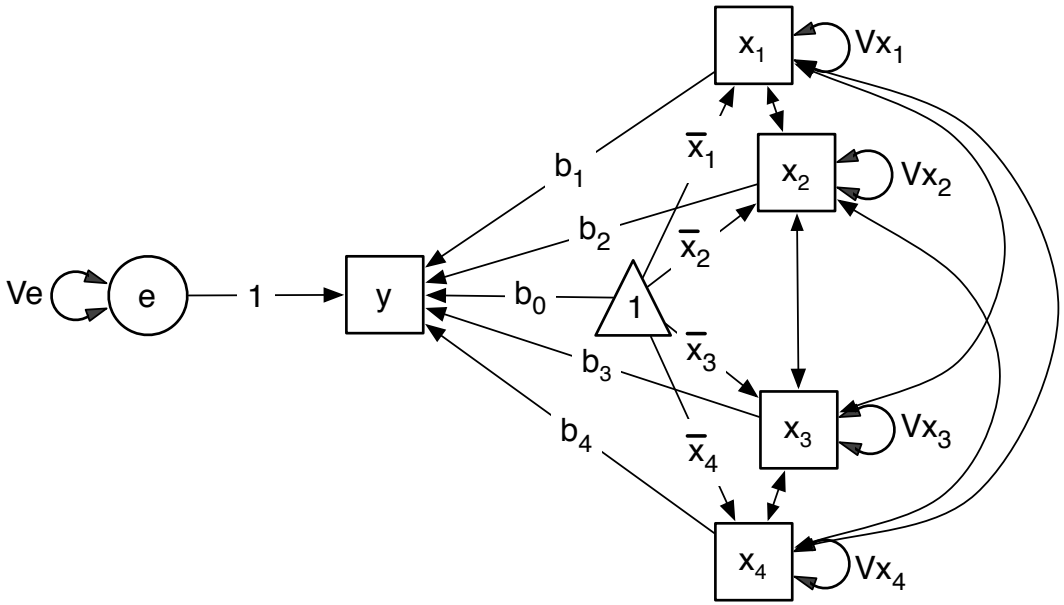
- Are $x1$ through $x4$ related to y?
- What are the fit statistics and degrees of freedom?
- Let's look at the model matrices that OpenMx created.
- Let's look more closely at the summary().

Standardized Multiple Regression



$$y_i = b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + e_i$$

FIML Multiple Regression



$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + e_i$$

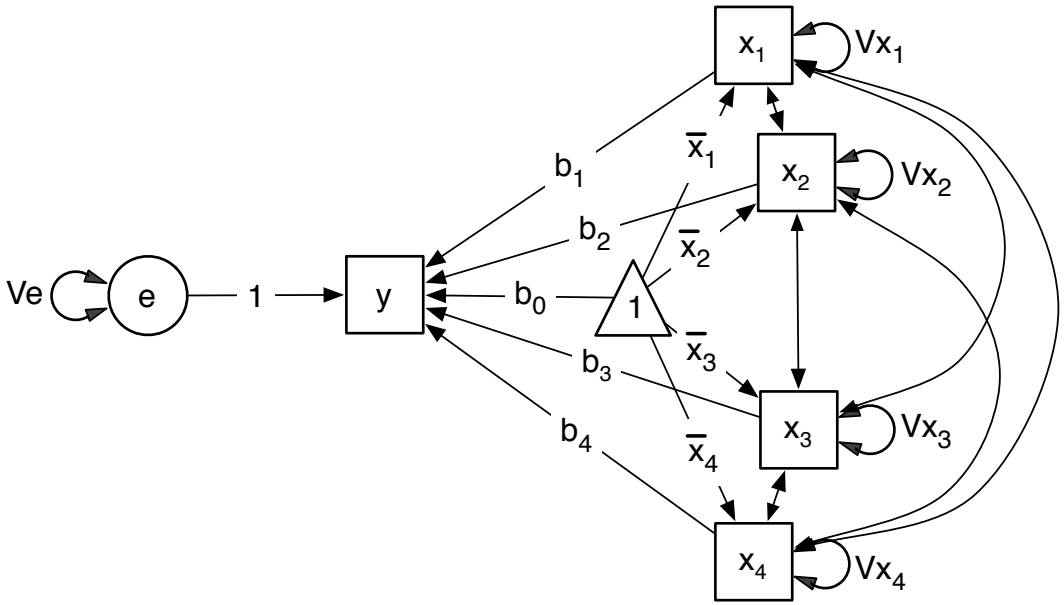
Running OpenMx

- The multiple regression FIML example is MultiRegRaw-OpenMx100214.R.

Evaluating the Multiple Regression FIML Result

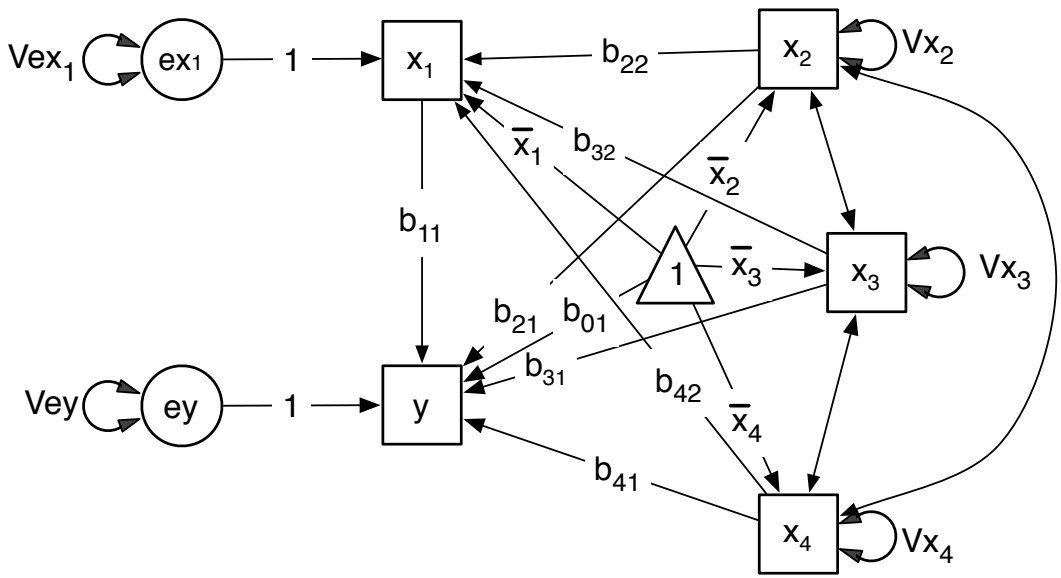
- Are $x1$ through $x4$ related to y?
- What are the fit statistics and degrees of freedom?
- Let's look at the model matrices that OpenMx created.
- Let's look more closely at the summary().
  - Now, what is the mean of y?

FIML Multiple Regression



$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i3} + b_4 x_{i4} + e_i$$

FIML Multivariate Regression



$$
\begin{aligned}
y_i &= b_{0y} + b_1 x_{i1} + b_{21} x_{i2} + b_{31} x_{i3} + b_{41} x_{i4} + e_{i1} \\
x_{i1} &= b_{0x1} + b_{22} x_{i2} + b_{32} x_{i3} + b_{42} x_{i4} + e_{i2}
\end{aligned}
$$

Running OpenMx

- The multiple regression FIML example is MultivariateRegRaw-OpenMx100214.R.

Evaluating the Multivariate Regression FIML Result

- Are $x1$ through $x4$ related to y?
- What are the fit statistics and degrees of freedom?
- Let's look at the model matrices that OpenMx created.
- Let's look more closely at the summary().
  - What is the mean of y?
  - What is the mean of $x1$?

Next Week

- Specifying OpenMx models using matrices.
- Regression and the General Linear Model using matrices.
- Review of Principal Components and Factor Analysis.
- Confirmatory Factor Analysis.
- Latent Variables.
- Read
  - Maruyama Chapter 7
  - McArdle (1990) Principles versus principals of structural factor analysis, *Multivariate Behavioral Research*, 25(1), 81-87.