The ‘simsem’ Vignette Template

# **Example 1: Getting Started**

0.7

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

0.7

0.7

0.7

0.7

0.7

0.51

1

1

0.5

0.51

0.51

0.51

0.51

0.51

# **Example 2: Covariance Matrix Specification**

1.0

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

0.6

0.7

1.0

1.1

0.9

1.1

0.8

0.9

0.4

0.5

0.8

0.4

0.4

0.8

*Y*7

*Y*8

*Y*9

1.0

1.2

1.1

0.4

0.8

0.5

0.6

0.3

0.2

# Example 3: Model Misspecification

1

*T*1

*T*2

*T*3

*T*4

1.2

1

0.25

*r* = 0.5

1.2

1.2

1.2

1

1

1

1 *± 0.1*

2 *± 0.1*

3

1

5

1

0

# Example 4: Random Parameters

*Y*1

*Y*2

*Y*3

*Y*4

1

1\*

1\*

1

*N*(0.3, 0.1)

*U*(-0.1, 0.1)

*U*(-0.1, 0.1)

*U*(0.3, 0.5)

*U*(0.3, 0.5)

*U*(0.5, 0.7)

1\* = Residual variance that results in total variance of 1

# Example 5: Equality Constraint

0.7

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

0.7

0.7

0.7

0.7

0.7

1

1

0.5

1\*

*Y*7

Equal

*U*(0.6, 0.8)

1\*

*Y*8

1\*

*x*

*x*

1\*

1\* = Residual variance that makes indicator variance of 1

*U*(0.3, 0.5)

*N*(0.6, 0.05)

*Trivial Misspecification*

1. All cross loadings have *U*(-0.2, 0.2)
2. All error correlations have *N*(0, 0.1)

1\*

1\*

1\*

1\*

1\*

## Syntax

## Remarks

# Example 6: Power Analysis in Model Evaluation

0.7

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

0.7

0.7

0.7

0.7

0.7

1\*

1

1

1\*

1\*

1\*

1\*

1\*

0.7

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

0.7

0.7

0.7

0.7

0.7

1\*

1

*U*(0.7, 0.9)

1\*

1\*

1\*

1\*

1\*

True Model

Serious Misspecification

1\* = Residual variance that makes indicator variance of 1

*Trivial Misspecification*:

1. All cross loadings have *U*(-0.2, 0.2), if applicable
2. All error correlations have *N*(0, 0.1)

# Example 7: Missing Data Handling

*U*(0.3, 0.6)

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

1

1\*

*Y*7

*Y*8

*Y*9

*U*(0.4, 0.9)

1

1

1

1\*

1\*

1\*

1\*

1\*

1\*

1\*

1\*

*N*(0.2, 0.1)

Constructs

Methods

*Trivial Misspecification*:

1. All cross loadings have *U*(-0.2, 0.2) only in the construct side.
2. All error correlations have *N*(0, 0.1)

1\* = Residual variance that makes indicator variance of 1

*U*(0.3, 0.6)

*U*(0.3, 0.6)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*N*(0.4, 0.1)

*N*(0.3, 0.1)

# Example 8: Planned Missing Design

*Trivial Misspecification*:

All cross loadings have *U*(-0.2, 0.2).

1\* = Residual variance that makes indicator variance of 1

*Y*25

*Y*36

*Y*37

*Y*48

*U*(0.4, 0.9)

1

1

1\*

1\*

1\*

1\*

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

…

…

*Y*1

*Y*12

*Y*13

*Y*24

1

1

1\*

*U*(0.4, 0.9)

*U*(0.1, 0.6)

…

…

1\*

1\*

1\*

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.1, 0.6)

*U*(0.1, 0.6)

*U*(0.1, 0.6)

*U*(0.1, 0.6)

# **Example 9: Nonnormal Distribution**

0.7

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

0.7

0.7

0.7

0.7

0.7

0.51

1

1

*U*(-0.5, 0.5)

0.51

0.51

0.51

0.51

0.51

*Y*7

*Y*8

0.51

0.51

*Y*9

*Y*10

*Y*11

*Y*12

1

0.51

0.51

0.51

0.51

0.7

0.7

*U*(-0.5, 0.5)

*U*(-0.5, 0.5)

0.7

0.7

0.7

0.7

*Trivial Misspecification*:

All cross loadings have *U*(-0.2, 0.2).

*Y*1 *~ t*(2)

*Y*2 *~ t*(3)

*Y*3 *~ t*(4)

*Y*5 *~ χ*2(3)

*Y*6 *~ χ*2(4)

*Y*7 *~ χ*2(5)

*Y*8 *~ χ*2(6)

*Y*9 *~ -χ*2(3)

*Y*10 *~ -χ*2(4)

*Y*11 *~ -χ*2(5)

*Y*12 *~ -χ*2(6)

*Y*4 *~ t*(5)

# **Example 10: Nonnormal Factor Distribution**

*Trivial Misspecification*:

1. All cross loadings have *U*(-0.3, 0.3).
2. All error correlations have *N*(0, 0.1).
3. All direct effects have *U*(-0.1, 0.1)

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

1

1

*U*(-0.5, 0.5)

1\*

1\*

1\*

*Y*7

*Y*8

*Y*9

1\*

*Y*10

*Y*11

*Y*12

1\*

1\*

1\*

1\*

1\*

1\*

1\*

1\*

1\*

1\*

*F*1 *~ χ*2(5)

*e*3 *~ N(0, 1)*

*e*4 *~ N(0, 1)*

*F*2 *~ χ*2(5)

*U*(0.3, 0.5)

*U*(0.3, 0.5)

*U*(0.5, 0.7)

*U*(-0.1, 0.1)

*U*(-0.1, 0.1)

1\* = Residual variance that makes indicator variance of 1

*U*(0.7, 0.9)

*U*(0.7, 0.9)

*U*(0.7, 0.9)

*U*(0.7, 0.9)

# **Example 11: Single Indicator**

*Trivial Misspecification*:

All error correlations have *N*(0, 0.1).

*Y*1

*Y*2

*Y*3

*Y*4

1

*U*(-0.5, 0.5)

0

*Y*5

1\*

1\* = Residual variance that makes indicator variance of 1

1

*U*(-0.5, 0.5)

*U*(-0.5, 0.5)

1\*

1\*

1\*

1

*U*(0.7, 0.9)

*U*(0.7, 0.9)

*U*(0.7, 0.9)

0

1

*F*1 *~ N*(0, 1)

*e*3 *~ N*(0, 1\*)

*F*2 *~ χ*2(5)

# **Example 12: Missing at Random and Auxiliary Variable**

*Missing*

Missing if

No missing if

*Y*1

*Y*2

*Y*3

1\*

1\*

1\*

*Y*4

*Y*5

*Y*6

1\*

1\*

1\*

1\*

*U*(0.3, 0.5)

1\* = Residual variance   
that makes indicator   
variance of 1

*U*(0.5, 0.7)

*U*(0.5, 0.7)

*Y*7

1

1\*

*U*(-0.4, 0.4)

*Trivial Misspecification*:

All cross loadings have *U*(-0.2, 0.2).

Overall = 10%

# **Example 13: Analyzing Real Data**

\*

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

\*

\*

\*

\*

\*

\*

1

1

\*

\*

\*

\*

\*

\*

*Y*7

*Y*8

*Y*9

\* = Estimated Parameters

\*

\*

1

\*

\*

\*

\*

\*

*Trivial Misspecification*:

All cross loadings have *U*(-0.2, 0.2).

\*

# **Example 14: Analyzing Real Data with Multiple Imputation**

*X*1

*X*2

*X*3

*Y*1

\*

*Y*2

\*

*Y*3

\*

*Y*4

\*

*Y*5

\*

*Y*6

\*

*Y*7

\*

*Y*8

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

*Trivial Misspecification*:

All cross loadings have *U*(-0.2, 0.2).

\* = Estimated Parameters

1

1

1

# **Example 15: Modeling a Covariate**

*Trivial Misspecification*:

All error correlations have *N*(0, 0.1).

*Y*1

*Y*2

*Y*3

1\*

1\*

1\*

*Y*4

*Y*5

*Y*6

1\*

1\*

1\*

1\*

*U*(0.3, 0.5)

1\* = Residual variance that makes indicator variance of 1

*U*(0.5, 0.7)

*U*(0.5, 0.7)

*Y*7

0

*U*(0.3, 0.5)

*U*(0.3, 0.5)

1

1

1

*Y*1

*Y*2

*Y*3

\*

\*

\*

*Y*4

*Y*5

*Y*6

1\*

\*

\*

\*

*Y*7

0

1

\*

1

\*

\*

\*

\*

\*

\*

\*

\*

\*

\* = Estimated Parameters

# **Example 16: Select a set of variables for analysis**

*Trivial Misspecification*:

All cross loadings have *U*(-0.2, 0.2) withi the same timepoint.

1\* = Residual variance that makes indicator variance of 1

1

1

1

1\* = *a*

1\* = *b*

1\* = *c*

*a*

*b*

*c*

Data Generation Model

*U*(0.3, 0.5) = *g*

*U*(0.3, 0.5) = *h*

*g*

*h*

*U*(0.5, 0.7) = *d*

*U*(0.5, 0.7) = *e*

*U*(0.5, 0.7) = *f*

*d*

*e*

*f*

*U*(0.3, 0.5)

*U*(0.3, 0.5)

*U*(0.3, 0.5)

1

1

Analysis Model

\*

\*

1

# **Example 17: Simulation with Varying Sample Size**

0.7

*Y*1

*Y*2

*Y*3

0.7

0.7

1\*

1

1

0.5

1\*

1\*

*Y*1

1\*

*Y*2

1\*

*Y*3

1\*

0.7

0.7

0.7

1\* = Residual variance that makes indicator variance of 1

# **Example 18: Simulation with Varying Sample Size and Percent Missing**

1

*Y*2

*Y*3

*Y*4

*Y*5

1.2

1

0.25

*r* = 0.5

1.2

1.2

1.2

1

1

1

1 *± 0.1*

2 *± 0.1*

3

1

5

1

*Y*1

0

1

1

*D ~ Bernoulli(0.5)*

0.5

0.1

# **Example 19: Simulation with Varying Sample Size and Parameters**

1\*

1\* = Residual variance that makes indicator variance of 1

*Y*1

*Y*2

*Y*3

1\*

1\*

1\*

*Y*4

*Y*5

1\*

1\*

*Y*6

*Y*7

*Y*8

1\*

1\*

1\*

*Y*9

*Y*10

1\*

1\*

1\*

*U*(0.3, 0.9)

*U*(0.3, 0.9)

*U*(0, 0.9)

# **Example 20: Continuous Sample Size in Comparing Models (Incomplete)**

## Model Description

0.7

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

0.7

0.7

0.7

0.7

0.7

1\*

1

1

1\*

1\*

1\*

1\*

1\*

0.7

*Y*1

*Y*2

*Y*3

*Y*4

*Y*5

*Y*6

0.7

0.7

0.7

0.7

0.7

1\*

1

*U*(0.7, 0.9)

1\*

1\*

1\*

1\*

1\*

True Model

Serious Misspecification

1\* = Residual variance that makes indicator variance of 1

*Trivial Misspecification*:

1. All cross loadings have *U*(-0.2, 0.2), if applicable
2. All error correlations have *N*(0, 0.1)

# **Example 21: Parceling Data in Simulation (Incomplete)**

## Model Description

*Trivial Misspecification*:

All cross loadings have *U*(-0.2, 0.2).

1\* = Residual variance that makes indicator variance of 1

*Y*25

*Y*36

*Y*37

*Y*48

*U*(0.4, 0.9)

1

1

1\*

1\*

1\*

1\*

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

…

…

*Y*1

*Y*12

*Y*13

*Y*24

1

1

1\*

*U*(0.4, 0.9)

*U*(0.1, 0.6)

…

…

1\*

1\*

1\*

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.4, 0.9)

*U*(0.1, 0.6)

*U*(0.1, 0.6)

*U*(0.1, 0.6)

*U*(0.1, 0.6)

Data Generation Model

Analysis Model

\*

*P*1

*P*2

*P*3

\*

\*

\*

1

1

\*

\*

*P*7

\*

*P*8

\*

*P*9

\*

\*

\*

\*

\*

*P*4

*P*5

*P*6

\*

\*

\*

1

1

\*

\*

*P*10

\*

*P*11

\*

*P*12

\*

\*

\*

\*

\*

\*

\*

\*

\*

\*

# Summary of Model Specification

Elementary Matrix Object

* Matrix Object
* Symmetric Matrix Object
* Vector Object

Analysis Set of Matrices Object

* CFA Object
* Path Analysis Object
* SEM Object

Data Object

Model Object

Result Object

Distribution Object

* Uniform Distribution
* Normal Distribution
* Etc.

Constraint Object

Misspecified Set of Matrices Object

* CFA Object
* Path Analysis Object
* SEM Object

Model Output Object

Data Output Object

Missing Object

Real Data

Data Distribution Object

Free Parameter Object

* CFA Object
* Path Analysis Object
* SEM Object

Function Object

Parameter Result Object

# List of Distribution Objects

| Distribution | Class | Constructor | Attributes |
| --- | --- | --- | --- |
| Beta | SimBeta | simBeta | shape1, shape2, ncp |
| Binomial | SimBinom | simBinom | size, prob |
| Cauchy | SimCauchy | simCauchy | location, scale |
| Chi-squared | SimChisq | simChisq | df, ncp |
| Exponential | SimExp | simExp | rate |
| F | SimF | simF | df1, df2, ncp |
| Gamma | SimGamma | simGamma | shape, rate |
| Geometric | SimGeom | simGeom | prob |
| Hypergeometric | SimHyper | simHyper | m, n, k |
| Log Normal | SimLnorm | simLnorm | meanlog, sdlog |
| Logistic | SimLogis | simLogis | location, scale |
| Negative Binomial | SimNbinom | simNbinom | size, prob |
| Normal | SimNorm | simNorm | mean, sd |
| Poisson | SimPois | simPois | lambda |
| t | SimT | simT | df, ncp |
| Uniform | SimUnif | simUnif | min, max |
| Weibull | SimWeibull | simWeibull | shape, scale |

# Fit Indices Details

The fit indices provided in the simsem package:

***Chi-square Test of the Target Model*** (). The value of chi-square is the -2 times log likelihood between the observed means and covariance matrix and model-implied means and covariance matrix. The degree of freedom (*dfT*) is the number of elements in the means and covariance matrix subtracted by the number of free parameters in the target model.

***Chi-square Test of the Baseline Model*** (). Mostly, the baseline model estimates means and variances of the observed data but not the covariances of the observed data. When there are auxiliary variables, the covariance of the auxiliary variables to all other variables (including themselves) are estimated. The chi-square value is the -2 times log likelihood between the observed means and covariance matrix and baseline model-implied means and covariance matrix. The degree of freedom (*dfB*) is the number of elements in means and covariance matrix and the number of free parameters in the baseline model.

***Comparative Fit Index*** (CFI). This index is one of the relative fit indices comparing between the fit of the target model and the fit of the baseline model. The minimum is 0 indicating bad fit and the maximum is 1 indicating perfect fit.

***Tucker-Lewis Index*** (TLI) or ***Non-Normed Fit Index*** (NNFI). This index is also one of the relative fit indices comparing between target model and baseline model. The minimum is 0 indicating bad fit and the maximum can be slightly greater than 1. The larger value indicates good fit.

***Akaike Information Criterion*** (AIC). This index is usually used to compare between two nonnested model. The model with smaller AIC provides better fit to the observed data.

where and is the number of free parameters

***Bayesian Information Criterion*** (BIC). This index is also usually used to compare between two nonnested model. The model with smaller BIC provides better fit to the observed data.

where *N* is sample size.

***Root Mean Squared Error of Approximation***(RMSEA). This index approximates the amount of misfit per degree of freedom. The minimum value is 0 indicating excellent fit.

***Standardized Root Mean Squared Residual***(SRMR). This index indicates the average discrepancy between observed correlations and model-implied correlations. The minimum value is 0 indicating excellent fit.

where is the observed covariance between indicator *i* and *j*, is the model-implied covariance between the indicator *i* and *j*, and *p* is the number of indicators.