Model Implied Instrumental Variable (MIIV) Methods using MIIVsem: An R Package for Structural Equation Models (SEMs)

Kenneth A Bollen

Zachary Fisher

University of North Carolina at Chapel Hill

INTRODUCTION

System Wide Maximum Likelihood (ML)

- ML estimator properties under ideal conditions
  - Consistent
  - Asymptotic unbiased
  - Asymptotic efficient
  - Asymptotic normality
  - Asymptotic standard errors

INTRODUCTION

System Wide Maximum Likelihood (ML)

- If your models are perfectly specified & your observed variables are from normal distributions, then
  - Go home, no need to attend workshop

INTRODUCTION

System Wide Maximum Likelihood (ML)

- If your models are *approximations* & your observed variables are from *nonnormal* distributions, then
  - You've come to right place
  - Approximations undermine ML properties
  - Bias & inconsistent estimator likely
  - Efficiency & accurate standard errors no longer guaranteed

INTRODUCTION

Other Issues with ML

- Underidentified models
  - Can prevent estimation & testing even if key equations in system are identified
- Nonconvergence
  - Prevent estimates from being obtained
  - Increasing iterations often does not help

INTRODUCTION

Other Issues with ML

- Spread bias from bad parts of model to good parts even in well specified equations
- Global tests of fit often significant
  - Not always easy to find source of problem
  - Bad measurement model?
  - Bad latent variable model?
  - Both?

INTRODUCTION

What do we need?

1. Estimator less likely to spread structural specification errors throughout system
2. Local estimates of equations
3. Local tests of equations
4. Ability to estimate identified equations, even if whole model not identified
5. Ideally a "distribution free" estimator
6. Noniterative without convergence problems

INTRODUCTION

Model Implied Instrumental Variables (MIIVs) addresses these:

1. MIIV-2SLS less likely to spread structural specification errors throughout system
2. Local estimates of equations
3. Local tests of equations
4. Ability to estimate identified equations, even if whole model not identified
5. A "distribution free" estimator
6. Noniterative so no convergence problems

INTRODUCTION

Purposes

1. Give an overview of MIIV-2SLS
2. Show you how to download the MIIVsem R package
3. Present and illustrate the primary steps in using MIIVsem
4. Teach you how to use & interpret MIIVsem input & output
5. Provide empirical examples that are estimated and tested with MIIVsem

Installing MIIVsem

First steps

- Install MIIVsem Version 0.5.8 from CRAN.
- Install lavaan
  - Will use for simulating data and estimating models later
- Type the following commands into the R console

```
install.packages("MIIVsem")
install.packages("lavaan")
```

Loading MIIVsem

Next, load MIIVsem

- Type the following command:

```
library("MIIVsem")
```

- Will load lavaan later when needed

Inputting Data File

First dataset is Bollen (1989)

- Comes with MIIVsem
- Accessed by typing `bollen1989a` in R console

To match the examples in the workshop we'll save the `bollen1989a` as `data` and rename the variables.

Inputting Data File

```
# save the political democracy
# dataset as data.
data <- bollen1989a

# rename the variables.
colnames(data) <- c("Z4","Z5","Z6",
                    "Z7","Z8","Z9",
                    "Z10","Z11","Z1",
                    "Z2", "Z3" )
```

PRIMARY INGREDIENTS

1. Specify Model
2. Transform Latent to Observed (L2O) variable model
3. Find Model Implied Instrumental Variables (MIIVs, pronounced to rhyme with "gives")
4. Estimate with Two Stage Least Squares (2SLS)
5. Test each overidentified equation

PRIMARY INGREDIENTS

1. Specify Model
   - Researcher lays out the latent variable and measurement models

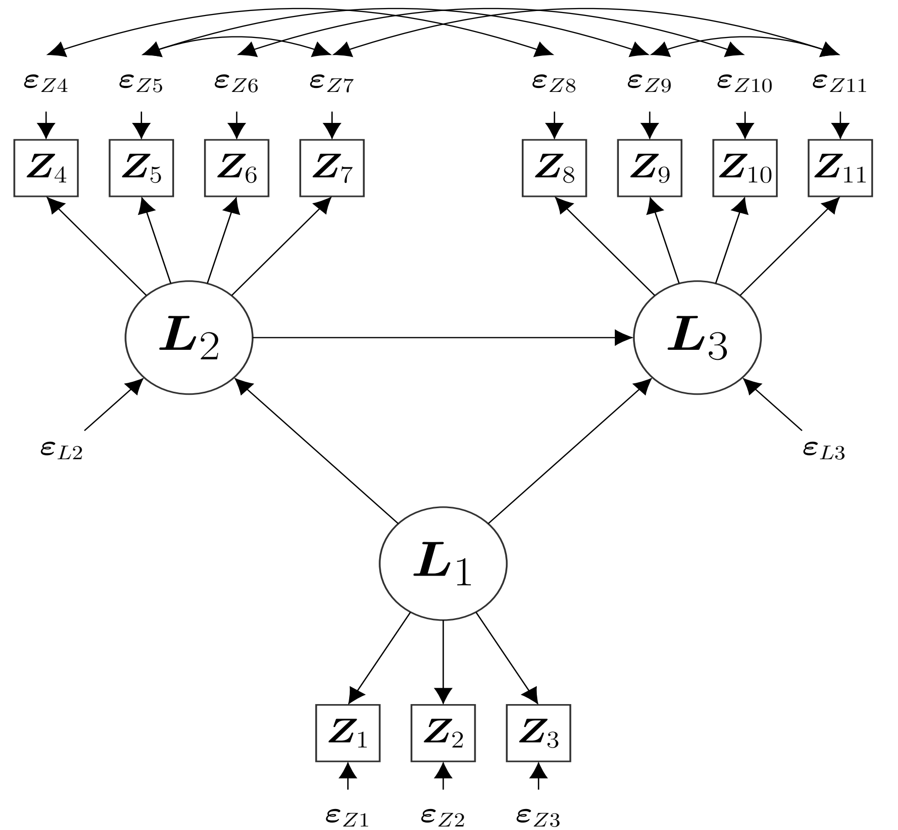
Industrialization and Political Democracy Example

L_1 = Industrialization at time 1

L_2 = Political Democracy at time 1

L_3 = Political Democracy at time 2

Z_1 to Z_{11} are indicators of L_1 to L_3



Industrialization and Political Democracy Example

Latent Variable Model

$$L_1 = \varepsilon_{L_1}$$

$$L_2 = \alpha_{L_2} + B_{21}L_1 + \varepsilon_{L_2}$$

$$L_3 = \alpha_{L_3} + B_{31}L_1 + B_{32}L_2 + \varepsilon_{L_3}$$

Measurement Model

$$Z_1 = L_1 + \varepsilon_{z1}$$

$$Z_2 = \alpha_{z2} + \Lambda_{21}L_1 + \varepsilon_{z2}$$

$$Z_3 = \alpha_{z3} + \Lambda_{31}L_1 + \varepsilon_{z3}$$

$$Z_4 = L_2 + \varepsilon_{z4}$$

$$Z_5 = \alpha_{z5} + \Lambda_{52}L_2 + \varepsilon_{z5}$$

$$Z_6 = \alpha_{z6} + \Lambda_{62}L_2 + \varepsilon_{z6}$$

$$Z_7 = \alpha_{z7} + \Lambda_{72}L_2 + \varepsilon_{z7}$$

$$Z_8 = L_3 + \varepsilon_{z8}$$

$$Z_9 = \alpha_{z9} + \Lambda_{93}L_3 + \varepsilon_{z9}$$

$$Z_{10} = \alpha_{z10} + \Lambda_{10,3}L_3 + \varepsilon_{z10}$$

$$Z_{11} = \alpha_{z11} + \Lambda_{11,3}L_3 + \varepsilon_{z11}$$

MIIVsem: Model Syntax

Main ingredient of MIIVsem:

- Model
- To define, for search or estimation, MIIVsem uses lavaan (Rosseel, 2012) syntax

Major operators to define relationships in model:

=~ "measured by" e.g., L1 =~ Z1

~ "regressed on" e.g., L5 ~ L4

~~ "covaries with" e.g., L2~~L3

* Assigns equality or numerical constraints

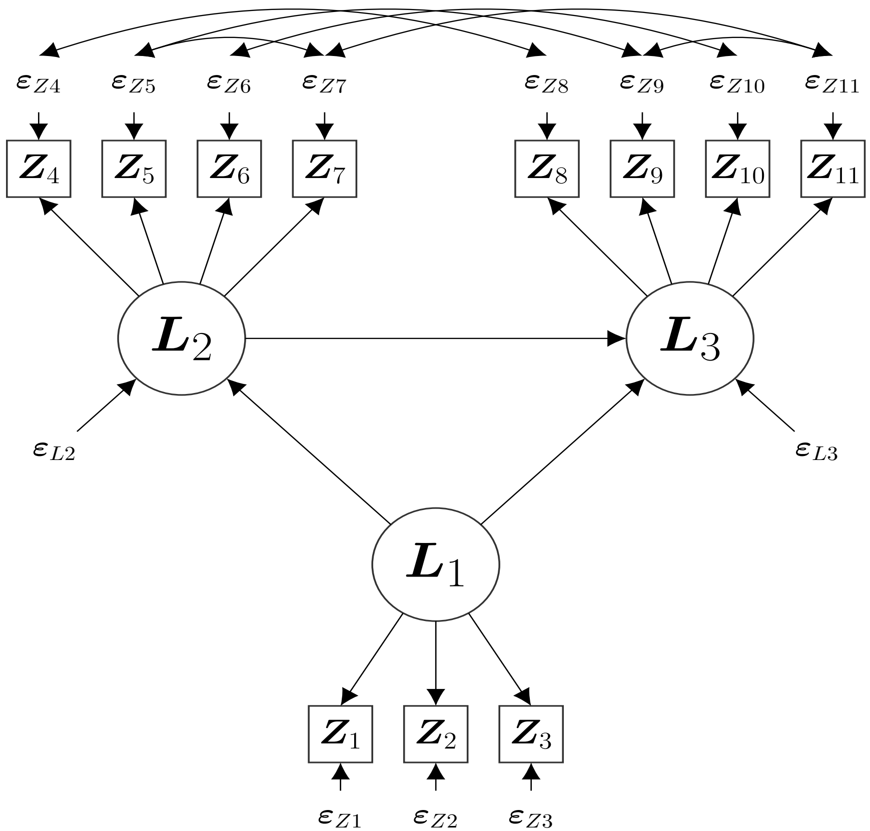
e.g., L4~ b*L3+b*L2 OR L1=~1*Z1+.5*Z2+Z3

MIIVsem: Industrialization-Democracy Example

In **R**, <- is used for assignment. Below we specify the industrialization-democracy model using our model syntax operators, surrounding the equations in single quotes. This creates an object in our workspace named model.indem1

```
model.indem1 <- '
  L1 =~ Z1 + Z2 + Z3
  L2 =~ Z4 + Z5 + Z6  + Z7
  L3 =~ Z8 + Z9 + Z10 + Z11

  L2  ~ L1
  L3  ~ L1 + L2

  Z4 ~~ Z8
  Z5 ~~ Z7 + Z9
  Z6 ~~ Z10
  Z7 ~~ Z11
  Z9 ~~ Z11 '
```

EXERCISE: In R, specify the model below.




```
model.indem1 <- '
  L1 =~ Z1 + Z2 + Z3
  L2 =~ Z4 + Z5 + Z6  + Z7
  L3 =~ Z8 + Z9 + Z10 + Z11

  L2  ~ L1
  L3  ~ L1 + L2

  Z4 ~~ Z8
  Z5 ~~ Z7 + Z9
  Z6 ~~ Z10
  Z7 ~~ Z11
  Z9 ~~ Z11 '
```

PRIMARY INGREDIENTS

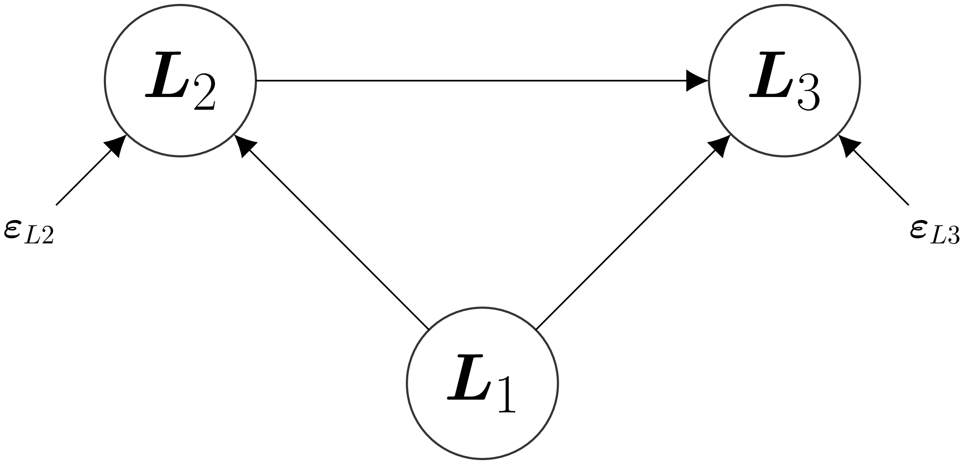
1. Specify Model ✓
2. Transform Latent to Observed (L2O) variable model
3. Find Model Implied Instrumental Variables (MIIVs, pronounced to rhyme with "gives")
4. Estimate with Two Stage Least Squares (2SLS)
5. Test each overidentified equation

PRIMARY INGREDIENTS

1. Specify Model ✓
2. Transform Latent to Observed (L2O) variable model (Bollen, 1996)
   - MIIVsem does this automatically
   - Here we illustrate how MIIVsem does it

PRIMARY INGREDIENTS

2. Transform Latent to Observed (L2O) variable model (Bollen, 1996)



PRIMARY INGREDIENTS

2. Transform Latent to Observed (L2O) variable model

$$L_1 = \varepsilon_{L_1}$$

$$L_2 = \alpha_{L_2} + B_{21}L_1 + \varepsilon_{L_2}$$

$$L_3 = \alpha_{L_3} + B_{31}L_1 + B_{32}L_2 + \varepsilon_{L_3}$$

$$Z_1 = L_1 + \varepsilon_{z1}$$

$$Z_4 = L_2 + \varepsilon_{z4}$$

$$Z_8 = L_3 + \varepsilon_{z8}$$

$$L_1 = Z_1 - \varepsilon_{z1}$$

$$L_2 = Z_4 - \varepsilon_{z4}$$

$$L_3 = Z_8 - \varepsilon_{z8}$$

PRIMARY INGREDIENTS

2. Transform Latent to Observed (L2O) variable model

Substitute scaling indicator minus error for each latent variable:

$$L_2 = \alpha_{L2} + B_{21}L_1 + \varepsilon_{L2} \quad \Rightarrow$$

$$\boxed{Z_4 = \alpha_{L2} + B_{21}Z_1 + u_4} \text{ with } u_4 = -B_{21}\varepsilon_{Z1} + \varepsilon_{Z4} + \varepsilon_{L2}$$

$$L_3 = \alpha_{L3} + B_{31}L_1 + B_{32}L_2 + \varepsilon_{L3} \quad \Rightarrow$$

$$\boxed{Z_8 = \alpha_{L3} + B_{31}Z_1 + B_{32}Z_4 + u_8} \text{ with } u_8 = -B_{31}\varepsilon_{Z1} - B_{32}\varepsilon_{Z4} + \varepsilon_{Z8} + \varepsilon_{L3}$$

Latent variable equations are transformed into
observed variable equations with composite errors.

Industrialization and Political Democracy Example

Latent Variable Model

$$L_1 = \varepsilon_{L_1}$$

$$L_2 = \alpha_{L_2} + B_{21}L_1 + \varepsilon_{L_2}$$

$$L_3 = \alpha_{L_3} + B_{31}L_1 + B_{32}L_2 + \varepsilon_{L_3}$$

Measurement Model

$$Z_1 = L_1 + \varepsilon_{z1}$$

$$Z_2 = \alpha_{z2} + \Lambda_{21}L_1 + \varepsilon_{z2}$$

$$Z_3 = \alpha_{z3} + \Lambda_{31}L_1 + \varepsilon_{z3}$$

$$Z_4 = L_2 + \varepsilon_{z4}$$

$$Z_5 = \alpha_{z5} + \Lambda_{52}L_2 + \varepsilon_{z5}$$

$$Z_6 = \alpha_{z6} + \Lambda_{62}L_2 + \varepsilon_{z6}$$

$$Z_7 = \alpha_{z7} + \Lambda_{72}L_2 + \varepsilon_{z7}$$

$$Z_8 = L_3 + \varepsilon_{z8}$$

$$Z_9 = \alpha_{z9} + \Lambda_{93}L_3 + \varepsilon_{z9}$$

$$Z_{10} = \alpha_{z10} + \Lambda_{10,3}L_3 + \varepsilon_{z10}$$

$$Z_{11} = \alpha_{z11} + \Lambda_{11,3}L_3 + \varepsilon_{z11}$$

EXERCISE: Do the L2O transformation for Z_5.

2. Transform Latent to Observed (L2O) variable model

Measurement Model

$$Z_1 = L_1 + \varepsilon_{z1}$$
$$Z_2 = \alpha_{z2} + \Lambda_{21} L_1 + \varepsilon_{z2}$$
$$Z_3 = \alpha_{z3} + \Lambda_{31} L_1 + \varepsilon_{z3}$$

$$Z_4 = L_2 + \varepsilon_{z4}$$
$$Z_5 = \alpha_{z5} + \Lambda_{52} L_2 + \varepsilon_{z5}$$
$$Z_6 = \alpha_{z6} + \Lambda_{62} L_2 + \varepsilon_{z6}$$
$$Z_7 = \alpha_{z7} + \Lambda_{72} L_2 + \varepsilon_{z7}$$

$$Z_8 = L_3 + \varepsilon_{z8}$$
$$Z_9 = \alpha_{z9} + \Lambda_{93} L_3 + \varepsilon_{z9}$$
$$Z_{10} = \alpha_{z10} + \Lambda_{10,3} L_3 + \varepsilon_{z10}$$
$$Z_{11} = \alpha_{z11} + \Lambda_{11,3} L_3 + \varepsilon_{z11}$$

Find L2O Transformation for: $Z_5 = \alpha_{z5} + \Lambda_{52} L_2 + \varepsilon_{z5}$

SOLUTION: Do the L2O transformation for Z_5.

2. Transform Latent to Observed (L2O) variable model

$$Z_4 = L_2 + \varepsilon_{z4} \Rightarrow L_2 = Z_4 - \varepsilon_{z4}$$

$$\boxed{Z_5 = \alpha_{z5} + \Lambda_{52} L_2 + \varepsilon_{z5}} \Rightarrow Z_5 = \alpha_{z5} + \Lambda_{52}(Z_4 - \varepsilon_{z4}) + \varepsilon_{z5}$$

$$Z_5 = \alpha_{z5} + \Lambda_{52} Z_4 - \Lambda_{52}\varepsilon_{z4} + \varepsilon_{z5} = \alpha_{z5} + \Lambda_{52} Z_4 + u_{z5}$$

SOLUTION: Do the L2O transformation for Z_5.

2. Transform Latent to Observed (L2O) variable model

$$Z_5 = \alpha_{z5} + \Lambda_{52} Z_4 + u_{z5}$$

Looks like simple regression with Z_5 dependent variable and Z_4 explanatory variable.

Why not use OLS?

Look at error,

$$Z_5 = \alpha_{z5} + \Lambda_{52} Z_4 - \Lambda_{52}\varepsilon_{z4} + \varepsilon_{z5}$$

$$C(Z_4, \varepsilon_{z4}) \neq 0$$

$$\Rightarrow OLS \text{ biased \& inconsistent}$$

PRIMARY INGREDIENTS

2. Transform Latent to Observed (L2O) variable model

Same problem for previous latent variable L2O transformation:

$$L_2 = \alpha_{L2} + B_{21}L_1 + \varepsilon_{L2} \quad \Rightarrow$$

$$\boxed{Z_4 = \alpha_{L2} + B_{21}Z_1 + u_4} \text{ with } u_4 = -B_{21}\varepsilon_{Z1} + \varepsilon_{Z4} + \varepsilon_{L2}$$

$$L_3 = \alpha_{L3} + B_{31}L_1 + B_{32}L_2 + \varepsilon_{L3} \quad \Rightarrow$$

$$\boxed{Z_8 = \alpha_{L3} + B_{31}Z_1 + B_{32}Z_4 + u_8} \text{ with } u_8 = -B_{31}\varepsilon_{Z1} - B_{32}\varepsilon_{Z4} + \varepsilon_{Z8} + \varepsilon_{L3}$$

PRIMARY INGREDIENTS

2. Transform Latent to Observed (L2O) variable model

$$Z_4 = \alpha_{L2} + B_{21} Z_1 + u_4 \text{ with } u_4 = -B_{21}\varepsilon_{Z1} + \varepsilon_{Z4} + \varepsilon_{L2}$$
$$Z_8 = \alpha_{L3} + B_{31} Z_1 + B_{32} Z_4 + u_8 \text{ with } u_8 = -B_{31}\varepsilon_{Z1} - B_{32}\varepsilon_{Z4} + \varepsilon_{Z8} + \varepsilon_{L3}$$

Problem: error correlates with Right Hand Side (RHS) Zs, OLS biased

Instrumental variables can help.

1. Correlate with RHS Zs
2. Not correlate with composite errors
3. At least as many instruments as RHS Zs

Finding suitable instruments is the next step in MIIV-2SLS.

PRIMARY INGREDIENTS

1. Specify Model ✓
2. Transform Latent to Observed (L2O) variable model ✓
3. Find Model Implied Instrumental Variables (MIIVs, pronounced to rhyme with "gives")
4. Estimate with Two Stage Least Squares (2SLS)
5. Tests each overidentified equation

PRIMARY INGREDIENTS

A Brief Digression: What are instrumental variables?

- If not OLS, what?
  - OLS failure due to error – explanatory variable correlation
- Instrumental variable estimators
  - Explain for simple regression

$$Y = \alpha + BX + \varepsilon$$

$$C(X, \varepsilon) \text{ not zero}$$

Figure 1

Path diagram of simple regression model with error-covariate (X) correlation.

PRIMARY INGREDIENTS

A Brief Digression: What are instrumental variables?

- Instrumental variable (IV) estimators
  - Suppose have variable V that:
    - Correlates with X
    - Uncorrelated with error

Figure 1

Path diagram of simple regression model with error-covariate (X) correlation.

$$C(Y,V) = C(\alpha + BX + \varepsilon, V)$$
$$= B\,C(X,V)$$

so we can find B by

$$B = \frac{C(Y,V)}{C(X,V)}$$

This is the Instrumental Variable (IV) estimator where sample covariances substituted for population covariances.

PRIMARY INGREDIENTS

3. Find Model Implied Instrumental Variables (MIIVs)
   - Key property of instruments is that they are uncorrelated with equation error
   - MIIV approach proposed in Bollen (1996) finds instruments among observed variables already part of model
     - If identified model, then MIIVs are generally part of model
     - No need to search outside of model
     - Structure of model implies which observed variables are uncorrelated with equation disturbance

PRIMARY INGREDIENTS

3. Find Model Implied Instrumental Variables (MIIVs)

General algorithm to find MIIVs (Bollen, 1996)

1. Focus on single equation
2. Find direct & indirect effects on the observed variables of each error in the composite error,
3. Eliminate the observed variables found in 2.,
4. Find the direct & indirect effects of any errors correlated with the composite error,
5. Eliminate the observed variables found in 4.,
6. Remaining observed variables are MIIVs.

PRIMARY INGREDIENTS

3. Find Model Implied Instrumental Variables (MIIVs)
   - MIIVsem finds MIIVs automatically
     - R: MIIVsem (Fisher, Bollen, Gates & Rönkkö)
   - Useful to illustrate process with examples

PRIMARY INGREDIENTS

3. Find Model Implied Instrumental Variables (MIIVs)

Consider first latent variable equation, latent political democracy (L_2) regressed on latent industrialization (L_1):

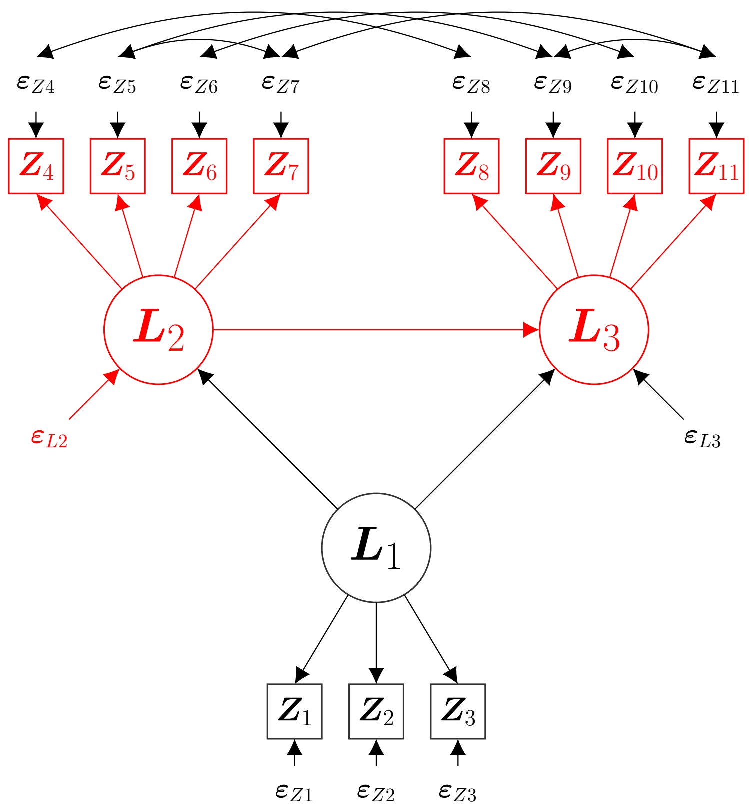
$$L_2 = \alpha_{L2} + B_{21}L_1 + \varepsilon_{L2} \quad \Rightarrow$$

$$\boxed{Z_4 = \alpha_{L2} + B_{21}Z_1 + u_4} \text{ with } u_4 = -B_{21}\varepsilon_{Z1} + \varepsilon_{Z4} + \varepsilon_{L2}$$

1. Find direct & indirect effects on observed variables of $\varepsilon_{Z1}, \varepsilon_{Z4}, \varepsilon_{L2}$.
   Let's start with ε_{L2} and return to path diagram of model.

PRIMARY INGREDIENTS

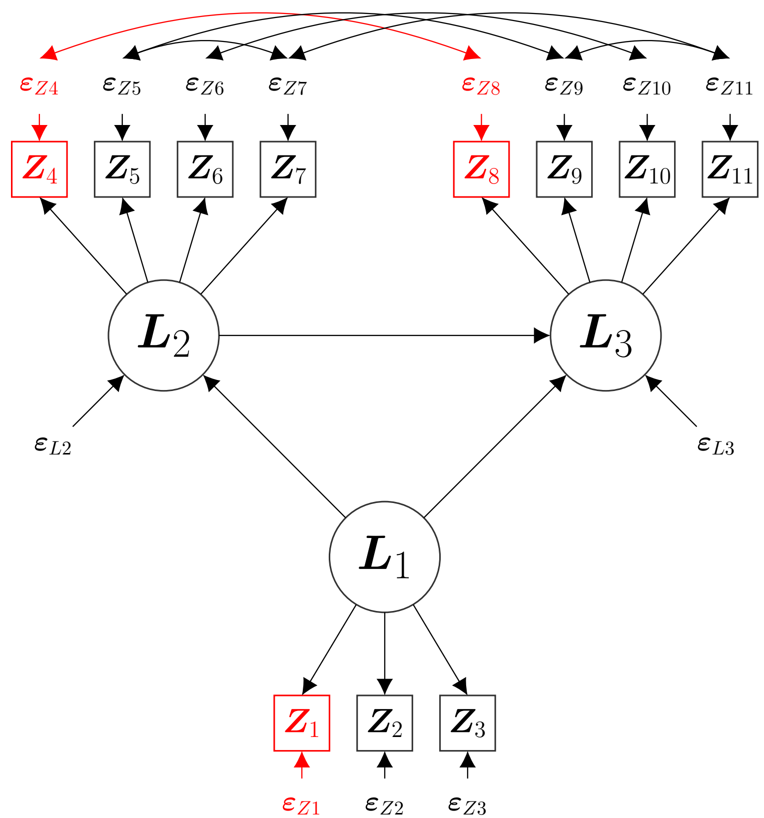
Find direct & indirect effects of ε_{L2}



Only variables NOT eliminated by ε_{L2} are Z_1, Z_2, Z_3.

PRIMARY INGREDIENTS

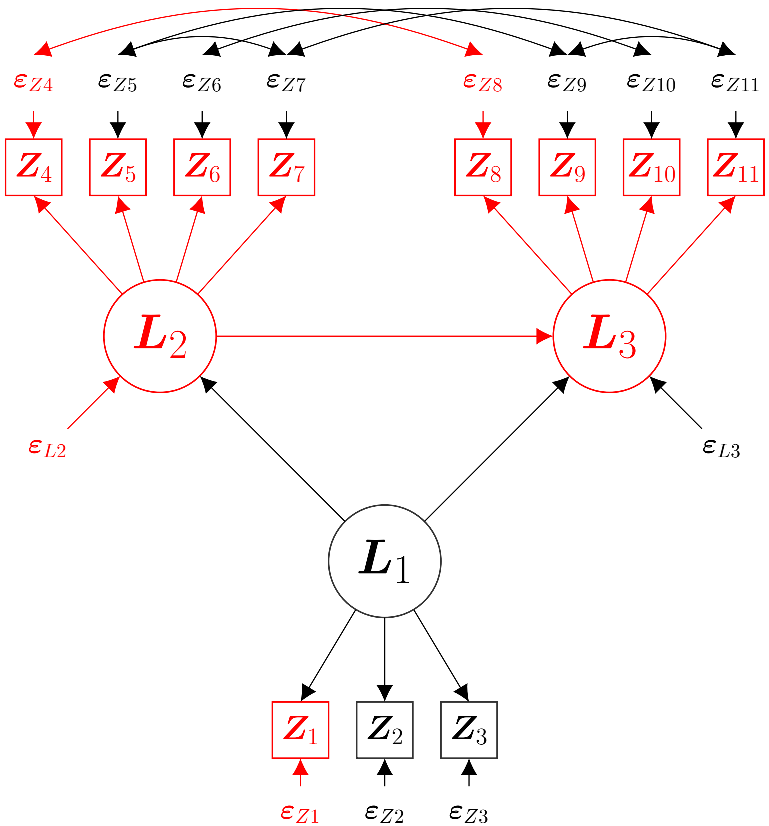
Find direct & indirect effects of $\varepsilon_{Z1}, \varepsilon_{Z4}$



Eliminates Z_1, Z_4, and Z_8 as MIIVs.

PRIMARY INGREDIENTS

Find direct & indirect effects of $\varepsilon_{Z1}, \varepsilon_{Z4}, \varepsilon_{L2}$



Z_2, Z_3 only MIIVs.

EXERCISE: Find the MIIVs for the Z_5 equation.

3. Find Model Implied Instrumental Variables (MIIVs)

- Return to previous L2O transformation

$$Z_4 = L_2 + \varepsilon_{z4} \Rightarrow L_2 = Z_4 - \varepsilon_{z4}$$

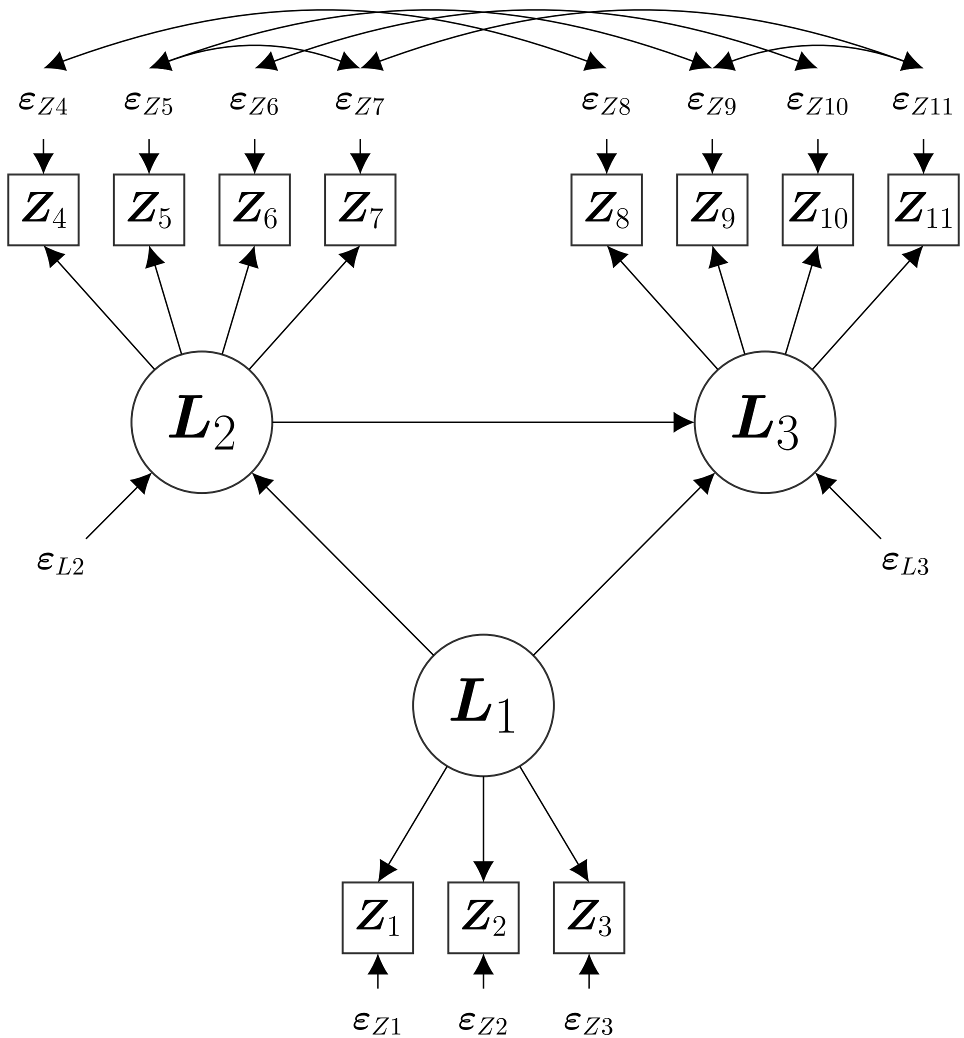
$$\boxed{Z_5 = \alpha_{z5} + \Lambda_{52} L_2 + \varepsilon_{z5}} \Rightarrow Z_5 = \alpha_{z5} + \Lambda_{52} Z_4 - \Lambda_{52} \varepsilon_{z4} + \varepsilon_{z5}$$

Find any variable directly or indirectly influenced by ε_{z4} or ε_{z5} and eliminate as MIIV.

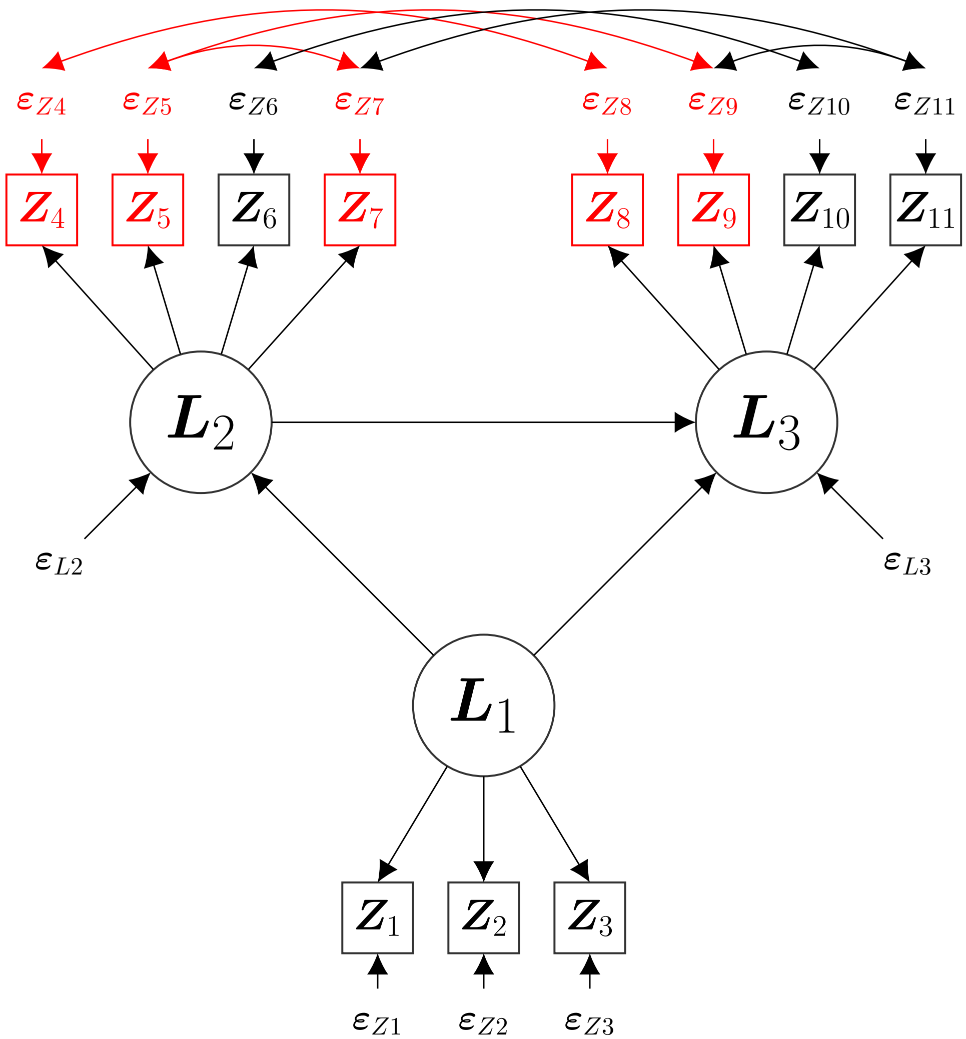
Find any errors that correlate with ε_{z4} or ε_{z5}

Eliminate as MIIVs any variable these influence.

EXERCISE: Find the MIIVs for the Z_5 equation.



SOLUTION: Find the MIIVs for the Z_5 equation.



SOLUTION: Find the MIIVs for the Z_5 equation.

3. Find Model Implied Instrumental Variables (MIIVs)

- Return to previous L2O transformation

$$Z_5 = \alpha_{z5} + \Lambda_{52} Z_4 - \Lambda_{52} \varepsilon_{z4} + \varepsilon_{z5}$$

MIIVs are:

Z_1 to Z_3, Z_6, Z_{10}, Z_{11}

The `miivs()`, or MIIV **s**earch, function in MIIVsem automatically performs the MIIV search.

The only argument needed to run `miivs()` is `model`.

```
model.indem1 <- '
        L1 =~ Z1 + Z2 + Z3
        L2 =~ Z4 + Z5 + Z6  + Z7
        L3 =~ Z8 + Z9 + Z10 + Z11

        L2  ~ L1
        L3  ~ L1 + L2

        Z4 ~~ Z8
        Z5 ~~ Z7 + Z9
        Z6 ~~ Z10
        Z7 ~~ Z11
        Z9 ~~ Z11  '

miivs(model.indem1)
```

EXERCISE: Use MIIVsem for the MIIV search.

The `miivs(), or` MIIV <u>**s**</u>earch, function in MIIVsem automatically performs the MIIV search.

```
model.indem1 <- '
    L1 =~ Z1 + Z2 + Z3
    L2 =~ Z4 + Z5 + Z6  + Z7
    L3 =~ Z8 + Z9 + Z10 + Z11

    L2  ~ L1
    L3  ~ L1 + L2

    Z4 ~~ Z8
    Z5 ~~ Z7 + Z9
    Z6 ~~ Z10
    Z7 ~~ Z11
    Z9 ~~ Z11  '

miivs(model.indem1)
```

The only argument needed to run `miivs()` is `model`.

Using the earlier defined model for industrialization-democracy run the MIIV search and check your output to see if it matches the output on the following slide.

Output:

```
## Model Equation Information
##
##  LHS  RHS     MIIVs
##  Z2   Z1      Z4, Z5, Z6, Z7, Z9, Z3, Z8, Z10, Z11
##  Z3   Z1      Z4, Z5, Z6, Z7, Z9, Z2, Z8, Z10, Z11
##  Z5   Z4      Z6, Z1, Z2, Z3, Z10, Z11
##  Z6   Z4      Z5, Z7, Z9, Z1, Z2, Z3, Z11
##  Z7   Z4      Z6, Z9, Z1, Z2, Z3, Z10
##  Z9   Z8      Z6, Z7, Z1, Z2, Z3, Z10
##  Z10  Z8      Z5, Z7, Z9, Z1, Z2, Z3, Z11
##  Z11  Z8      Z5, Z6, Z1, Z2, Z3, Z10
##  Z4   Z1      Z2, Z3
##  Z8   Z1, Z4  Z5, Z6, Z7, Z2, Z3
```

Note: The `summary` method for `miivs` objects provides additional options for displaying the MIIV search information:

```
summary(miivs(model.indem1), eq.info = TRUE)
```

PRIMARY INGREDIENTS

1. Specify Model ✓
2. Transform Latent to Observed (L2O) variable model ✓
3. Find Model Implied Instrumental Variables (MIIVs) ✓
4. Estimate with Two Stage Least Squares (2SLS)
5. Tests each overidentified equation

PRIMARY INGREDIENTS

4. Estimate with Two Stage Least Squares (2SLS)

In general,

$\mathbf{Y}_j$ = vector containing values of *j*th dependent variable for L2O equation

$\mathbf{Z}_j$ = matrix of explanatory variables on RHS of same *j*th L2O equation

$\mathbf{V}_j$ = matrix of MIIVs for same *j*th L2O equation

2SLS estimator of coefficents is $\left(\hat{\mathbf{Z}}_j'\hat{\mathbf{Z}}_j\right)^{-1}\hat{\mathbf{Z}}_j'\mathbf{Y}_j$

where $\hat{\mathbf{Z}}_j = \mathbf{V}_j\left(\mathbf{V}_j'\mathbf{V}_j\right)^{-1}\mathbf{V}'\mathbf{Z}_j$

Noniterative

No issues with convergence

PRIMARY INGREDIENTS

4. Estimate with Two Stage Least Squares (2SLS)

Consider first latent variable equation, latent political democracy (L_2) regressed on latent industrialization (L_1):

$$L_2 = \alpha_{L2} + B_{21}L_1 + \varepsilon_{L2} \Rightarrow \boxed{Z_4 = \alpha_{L2} + B_{21}Z_1 + u_4} \quad \text{MIIVs are: } Z_2, Z_3$$

$$\mathbf{Y}_j = \begin{bmatrix} Z_{41} \\ Z_{42} \\ \vdots \\ Z_{4N} \end{bmatrix} \qquad \mathbf{Z}_j = \begin{bmatrix} 1 & Z_{11} \\ 1 & Z_{12} \\ \vdots & \vdots \\ 1 & Z_{1N} \end{bmatrix} \qquad \mathbf{V}_j = \begin{bmatrix} 1 & Z_{21} & Z_{31} \\ 1 & Z_{22} & Z_{32} \\ \vdots & \vdots & \vdots \\ 1 & Z_{2N} & Z_{3N} \end{bmatrix}$$

2SLS estimator of coefficents is $\left(\hat{\mathbf{Z}}_j'\hat{\mathbf{Z}}_j\right)^{-1}\hat{\mathbf{Z}}_j'\mathbf{Y}_j$

where $\hat{\mathbf{Z}}_j = \mathbf{V}_j\left(\mathbf{V}_j'\mathbf{V}_j\right)^{-1}\mathbf{V}'\mathbf{Z}_j$

PRIMARY INGREDIENTS

4. Estimate with Two Stage Least Squares (2SLS)

Comparison	MIIV-2SLS	ML
Consistency	✓	✓
Asymp. unbiased	✓	✓
Asymp. normal	✓	✓
Asymp. efficient	✓*	✓
Asymp. s.e.	✓	✓
Noniterative	✓	-
Nonnormal robust	✓	-**
No SEM software needed	✓	-
Overidentification test	equation	model

*2SLS efficient among limited information estimators.

**Corrected significance tests available.

PRIMARY INGREDIENTS

4. Estimate with Two Stage Least Squares (2SLS)

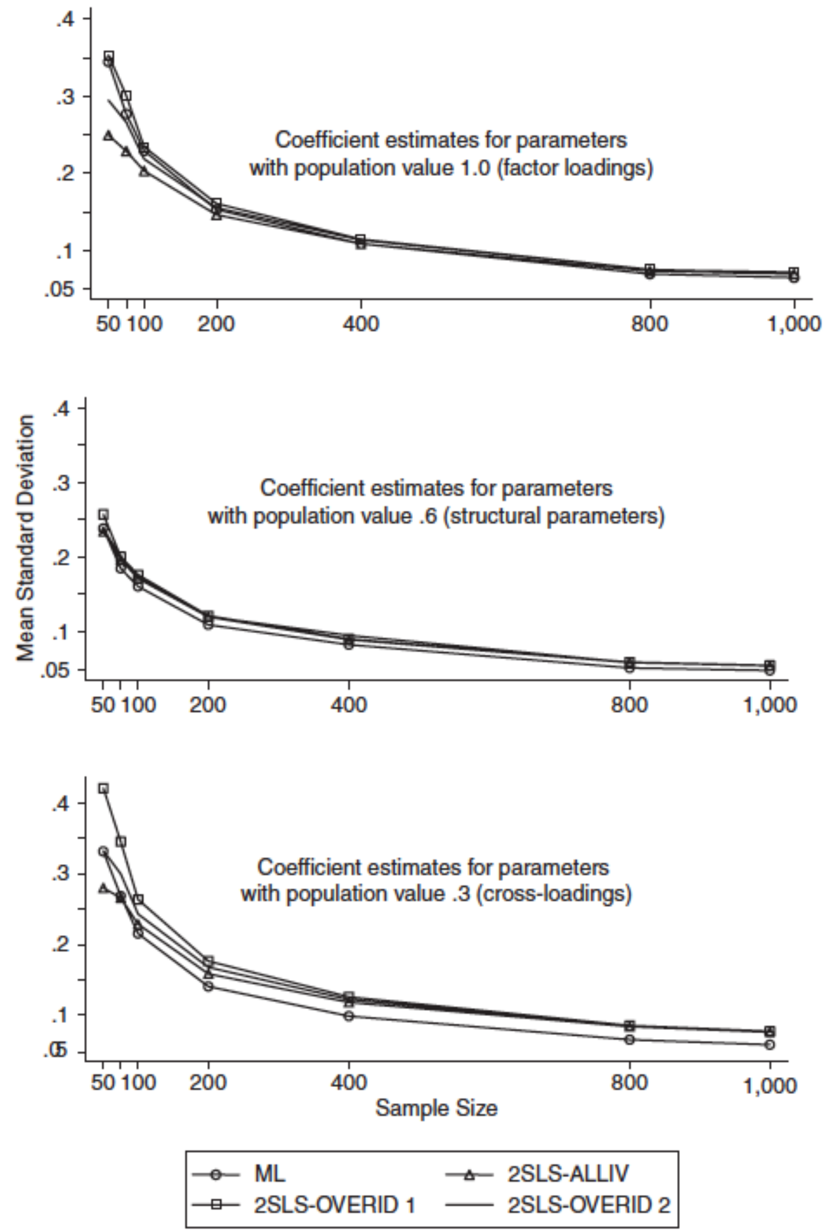
Illustration of ML and MIIV-2SLS simulation Bollen, Kirby, Curran, Paxton, & Chen (2007)

Graph on next page gives standard deviation of parameters under ideal conditions for ML:

Normality

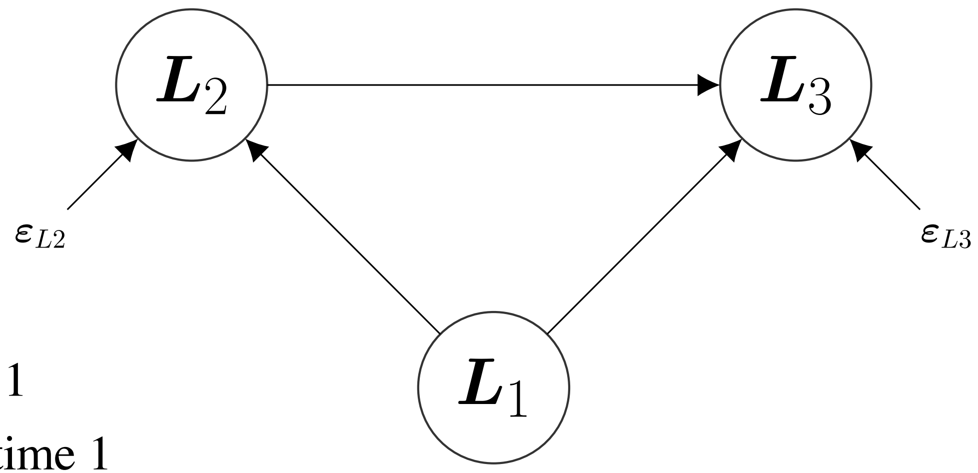
Correct specification

Mean Standard Deviation of Estimates From Four Estimators by Sample Size for Parameter Estimates From Specification 1, the Correctly Specified Model



PRIMARY INGREDIENTS

4. Estimate with Two Stage Least Squares (2SLS)
   - Return to latent variable model for example



L_1 = Industrialization at time 1

L_2 = Political Democracy at time 1

L_3 = Political Democracy at time 2

PRIMARY INGREDIENTS

`miive()`

- MIIV **e**stimation function in MIIVsem
- Use miive() to estimate the industrialization-democracy example using MIIV-2SLS.
- "data" is name assigned to the industrialization-democracy example
- Next slide recaps all the steps for estimation for industrialization-democracy example.
  - Already loaded MIIVsem, renamed the data and specified the model,
  - Only step left is to run miive.

EXERCISE: Estimate the Industrialization-Democracy model.

```
model.indem1 <- '
        L1 =~ Z1 + Z2 + Z3
        L2 =~ Z4 + Z5 + Z6  + Z7
        L3 =~ Z8 + Z9 + Z10 + Z11

        L2  ~ L1
        L3  ~ L1 + L2

        Z4 ~~ Z8
        Z5 ~~ Z7 + Z9
        Z6 ~~ Z10
        Z7 ~~ Z11
        Z9 ~~ Z11  '

data <- bollen1989a
colnames(data) <- c("Z4", "Z5", "Z6", "Z7",
                    "Z8", "Z9", "Z10", "Z11",
                    "Z1", "Z2", "Z3")
miive(model.indem1, data)
```

MIIVsem: Header for estimation output.

Number of observations: *the number of observations used for all equations in the system*.

Number of equations: *the total number of L2O equations estimated*.

Estimator: *the estimator used, either MIIV-2SLS or PIV*.

```
## MIIVsem (0.5.8) results
##
## Number of observations                              75
## Number of equations                                 10
## Estimator                                    MIIV-2SLS
## Standard Errors                               standard
## Missing                                       listwise
```

MIIVsem: Header for estimation output.

Standard errors: *the method used to compute standard errors, this will be discussed again when we discuss the bootstrap*.

Missing: *listwise is the default. In our dataset there was no missing data, this can be verified by looking at the number of observations.*

```
## MIIVsem (0.5.8) results
##
## Number of observations                                      75
## Number of equations                                         10
## Estimator                                            MIIV-2SLS
## Standard Errors                                       standard
## Missing                                               listwise
```

PRIMARY INGREDIENTS

```
## Parameter Estimates:
##
##
## STRUCTURAL COEFFICIENTS:
##                    Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
##   L1 =~
##     Z1                1.000
##     Z2                2.078    0.128   16.171    0.000    8.301    8    0.405
##     Z3                1.751    0.149   11.782    0.000    8.738    8    0.365
##   L2 =~
##     Z4                1.000
##     Z5                1.139    0.179    6.371    0.000    8.409    5    0.135
##     Z6                0.969    0.140    6.924    0.000    5.874    6    0.437
##     Z7                1.210    0.139    8.713    0.000    4.276    5    0.510
##   L3 =~
##     Z8                1.000
##     Z9                1.051    0.165    6.377    0.000    8.712    5    0.121
##     Z10               1.180    0.151    7.814    0.000    9.538    6    0.146
##     Z11               1.203    0.154    7.798    0.000    2.795    5    0.731
##
##   L2 ~
##     L1                1.261    0.426    2.962    0.003    0.503    1    0.478
##   L3 ~
##     L1                1.123    0.312    3.598    0.000    0.801    3    0.849
##     L2                0.724    0.101    7.140    0.000
```

PRIMARY INGREDIENTS

```
## INTERCEPTS:
##                       Estimate  Std.Err  z-value  P(>|z|)
##     L2                  -0.909    2.170   -0.419    0.675
##     L3                  -4.499    1.424   -3.160    0.002
##     Z1                   0.000
##     Z10                  0.135    0.830    0.163    0.870
##     Z11                 -2.137    0.853   -2.505    0.012
##     Z2                  -5.711    0.654   -8.727    0.000
##     Z3                  -5.292    0.758   -6.985    0.000
##     Z4                   0.000
##     Z5                  -1.969    1.044   -1.886    0.059
##     Z6                   1.265    0.814    1.553    0.120
##     Z7                  -2.160    0.814   -2.654    0.008
##     Z8                   0.000
##     Z9                  -2.418    0.909   -2.659    0.008
```

PRIMARY INGREDIENTS

1. Specify Model ✓
2. Transform Latent to Observed (L2O) variable model ✓
3. Find Model Implied Instrumental Variables (MIIVs) ✓
4. Estimate w/ Two Stage Least Squares (2SLS) ✓
5. Tests each overidentified equation

PRIMARY INGREDIENTS

5. Tests each overidentified equation

$$\frac{\hat{\mathbf{u}}\mathbf{V}(\mathbf{V}'\mathbf{V})^{-1}\mathbf{V}'\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}/\mathrm{N}} \overset{a}{\sim} \chi^2$$

where

$\hat{\mathbf{u}} =$ 2SLS residuals

$\mathbf{V} =$ MIIVs

N = sample size

df = # MIIVs - # endogenous regressors

PRIMARY INGREDIENTS

5. Tests each overidentified equation

Sargan Test:

H_0: MIIVs uncorrelated with equation error

H_a: At least 1 MIIV correlates with error

Reject H_0 is evidence against model because model led to MIIVs.

PRIMARY INGREDIENTS

```
## Parameter Estimates:
##
##
## STRUCTURAL COEFFICIENTS:
##                       Estimate  Std.Err  z-value  P(>|z|)   Sargan   df    P(Chi)
##   L1 =~
##     Z1                   1.000
##     Z2                   2.078    0.128   16.171    0.000    8.301    8     0.405
##     Z3                   1.751    0.149   11.782    0.000    8.738    8     0.365
##   L2 =~
##     Z4                   1.000
##     Z5                   1.139    0.179    6.371    0.000    8.409    5     0.135
##     Z6                   0.969    0.140    6.924    0.000    5.874    6     0.437
##     Z7                   1.210    0.139    8.713    0.000    4.276    5     0.510
##   L3 =~
##     Z8                   1.000
##     Z9                   1.051    0.165    6.377    0.000    8.712    5     0.121
##     Z10                  1.180    0.151    7.814    0.000    9.538    6     0.146
##     Z11                  1.203    0.154    7.798    0.000    2.795    5     0.731
##
##   L2 ~
##     L1                   1.261    0.426    2.962    0.003    0.503    1     0.478
##   L3 ~
##     L1                   1.123    0.312    3.598    0.000    0.801    3     0.849
##     L2                   0.724    0.101    7.140    0.000
```

EXERCISE: Adjust Sargan Test for multiple comparisons.

Multiple testing problem

- `sarg.adjust` argument of the `miive()`
  - p-value adjustment method for the Sargan test.

Reestimate the industrialization-democracy example

- Use Holm ("holm") correction for multiple comparisons
- Compare p-values to unadjusted counterparts

```
miive(model.indem1, data, sarg.adjust = "holm")
```

PRIMARY INGREDIENTS

```
## Parameter Estimates:
##
##
## STRUCTURAL COEFFICIENTS:
##                    Estimate  Std.Err  z-value  P(>|z|)   Sargan   df    P(Chi)
##   L1 =~
##     Z1                1.000
##     Z2                2.078    0.128   16.171    0.000    8.301    8     1.000
##     Z3                1.751    0.149   11.782    0.000    8.738    8     1.000
##   L2 =~
##     Z4                1.000
##     Z5                1.139    0.179    6.371    0.000    8.409    5     1.000
##     Z6                0.969    0.140    6.924    0.000    5.874    6     1.000
##     Z7                1.210    0.139    8.713    0.000    4.276    5     1.000
##   L3 =~
##     Z8                1.000
##     Z9                1.051    0.165    6.377    0.000    8.712    5     1.000
##     Z10               1.180    0.151    7.814    0.000    9.538    6     1.000
##     Z11               1.203    0.154    7.798    0.000    2.795    5     1.000
##
##   L2 ~
##     L1                1.261    0.426    2.962    0.003    0.503    1     1.000
##   L3 ~
##     L1                1.123    0.312    3.598    0.000    0.801    3     1.000
##     L2                0.724    0.101    7.140    0.000
```

PRIMARY INGREDIENTS

1. Specify Model
2. Transform Latent to Observed (L2O) variable model
3. Find Model Implied Instrumental Variables (MIIVs, pronounced to rhyme with "gives")
4. Estimate with Two Stage Least Squares (2SLS)
5. Test each overidentified equation

ROBUSTNESS

1. Distributional robustness
   - Properties of MIIV-2SLS are "distribution-free"
   - Asymptotic, but do not assume normal error or observed variables
   - Bootstrap option in MIIVsem permits alternative way to estimate standard errors of parameter estimates

MIIVsem: Bootstrap SEs and CIs.

MIIVsem bootstrap standard errors

- Requested by setting `se = "boot"`
- Default number of replications is 1000
- Adjusted using the bootstrap argument (e.g. `bootstrap = 500`)

Reestimate the industrialization-democracy model with bootstrap standard errors

- Also set the `boot.ci` argument to "`perc`" which requests a confidence interval from the empirical 2.5th and 97.5th percentiles of the bootstrap sample.

```
miive(model.indem1, data, se = "boot", boot.ci = "perc")
```

MIIVsem: Bootstrap SEs and CIs.

Header information

- Shows boostrap standard errors requested
- Lists number of replications requested
- Includes number of successful replications
- Documents confidence interval method

```
MIIVsem (0.5.2) results

Number of observations                                   75
Number of equations                                      10
Estimator                                         MIIV-2SLS
Standard Errors                                   bootstrap
Missing                                            listwise

Bootstrap reps requested                               1000
Bootstrap reps successful                              1000
Bootstrap intervals                              Percentile
```

Note: Additional options for constructing bootstrap CIs are available. See the `boot.ci` argument of `miive` for more choices.

MIIVsem: Bootstrap SEs and CIs.

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err    Lower    Upper   Sargan    df   P(Chi)
  L1 =~
    Z1                1.000
    Z2                2.078    0.127    1.787    2.300    8.301     8    0.405
    Z3                1.751    0.132    1.461    1.972    8.738     8    0.365
  L2 =~
    Z4                1.000
    Z5                1.139    0.134    0.850    1.373    8.409     5    0.135
    Z6                0.969    0.137    0.696    1.237    5.874     6    0.437
    Z7                1.210    0.122    0.950    1.434    4.276     5    0.510
  L3 =~
    Z8                1.000
    Z9                1.051    0.142    0.711    1.303    8.712     5    0.121
    Z10               1.180    0.148    0.858    1.452    9.538     6    0.146
    Z11               1.203    0.148    0.891    1.484    2.795     5    0.731

  L2 ~
    L1                1.261    0.422    0.406    2.059    0.503     1    0.478
  L3 ~
    L1                1.123    0.297    0.524    1.745    0.801     3    0.849
    L2                0.724    0.093    0.540    0.904
```

EXERCISE: Bootstrap SEs and CIs.

Reestimate the political democracy example with bootstrap standard errors. Try setting the number of replications using bootstrap to a different number and compare the results to those obtained from 1000 bootstrap replications.

For example, below we request 500 bootstrap replications.

```
miive(model.indem1, data, se = "boot",
      boot.ci = "perc", bootstrap = 500)
```

EXERCISE: Bootstrap SEs and CIs.

Below is code for combining estimates from four fitted models for easy comparison.

```
# Fit Models
fit.standard  <- miive(model.indem1, data)
fit.boot.250  <- miive(model.indem1, data, se = "boot", bootstrap = 250)
fit.boot.500  <- miive(model.indem1, data, se = "boot", bootstrap = 500)
fit.boot.1000 <- miive(model.indem1, data, se = "boot", bootstrap = 1000)

# Save Estimates
est.standard  <- estimatesTable(fit.standard)[, c("lhs", "op", "rhs", "se")]
est.boot.250  <- estimatesTable(fit.boot.250)[, c("lhs", "op", "rhs", "se")]
est.boot.500  <- estimatesTable(fit.boot.500)[, c("lhs", "op", "rhs", "se")]
est.boot.1000 <- estimatesTable(fit.boot.1000)[,c("lhs", "op", "rhs", "se")]

list.est   <- list(est.standard,est.boot.250,est.boot.500, est.boot.1000)
compare.se <- Reduce(function(...) merge(..., by= c("lhs", "op", "rhs")), list.est)

colnames(compare.se) <- c("","","", "standard", "boot.250", "boot.500", "boot.1000")

compare.se
```

SOLUTION: Bootstrap SEs and CIs.

Comparison of standard errors.

			standard	boot.250	boot.500	boot.1000
L1	=~	Z1	0.0000000	0.00000000	0.00000000	0.00000000
L1	=~	Z2	0.1284989	0.13149889	0.12511317	0.13265177
L1	=~	Z3	0.1486077	0.13369027	0.12816840	0.12758051
L2	=~	Z4	0.0000000	0.00000000	0.00000000	0.00000000
L2	=~	Z5	0.1788162	0.13971253	0.13728272	0.13531398
L2	=~	Z6	0.1400277	0.12746333	0.13526307	0.13095981
L2	=~	Z7	0.1388712	0.11816898	0.11031515	0.11557476
L2	~	L1	0.4257016	0.40698842	0.41878643	0.40293661
L3	=~	Z10	0.1510235	0.15270889	0.14275581	0.14965067
L3	=~	Z11	0.1542894	0.15037637	0.13776855	0.15206394
L3	=~	Z8	0.0000000	0.00000000	0.00000000	0.00000000
L3	=~	Z9	0.1647411	0.14737877	0.14925931	0.13995522
L3	~	L1	0.3121788	0.29706550	0.29238986	0.31008589
L3	~	L2	0.1014416	0.09298976	0.09572984	0.09957574

ROBUSTNESS

Table 5

MIIV-2SLS Robustness Conditions for Structural Misspecifications (Bollen, 2020b; Bollen et al., 2018)[a]

1. The MIIV-2SLS estimator of the coefficients of the measurement model is robust-unchanged to structural misspecifications in the latent variable model.
2. The MIIV-2SLS estimator of the coefficients of the latent variable model is robust-unchanged to having the incorrect covariance matrices among the unique factors (errors) of the measurement model for the nonscaling indicators.
3. The MIIV-2SLS estimator of the coefficients of the latent variable model is robust-unchanged to omitted cross-loadings as long as the omitted nonzero cross-loadings when present do not create indirect paths from the latent variable equation error to one or more of the MIIVs from the misspecified model.
4. The MIIV-2SLS estimator of the coefficients of the latent variable model is robust-unchanged to the omission of correlated unique factors (errors) for the scaling indicators of latent variables that are included in the true latent variable equation.
5. The MIIV-2SLS estimator of the coefficients of the latent variable model is *not* robust-unchanged to omitted cross-loadings if the omitted cross-loadings when present create nonzero indirect paths from the latent variable equation error to one or more of the MIIVs from the misspecified model.
6. The MIIV-2SLS estimator is robust-unchanged to the omission of any observed variable that is not part of the L2O transformed equation and is not among the MIIVs for that equation.
7. The MIIV-2SLS estimator of factor loadings in an indicator equation is robust-unchanged when the indicator equation is correct and none of the errors in the L2O version of the indicator equation correlate with other indicators in the model.
8. The MIIV-2SLS estimator of an indicator equation is robust-unchanged under the conditions given in #7. even when among the remaining indicators there are: (a) omitted correlated errors, (b) omitted crossloadings, (c) causal indicators treated as reflective, or (d) direct effects between these indicators.
9. The MIIV-2SLS estimator of the coefficients of an indicator equation is *not* robust-unchanged to: (a) omitted correlated errors involving the indicator or scaling indicator with the errors of other indicators, (b) omitted crossloadings that directly affect the indicator or scaling indicator, (c) mistakenly treating the indicator or scaling indicator as reflective when one or both are causal indicators, or (d) omitting direct effects between the scaling or indicator of interest and one or more remaining indicators.

Note. MIIV-2SLS = Model implied instrumental variable, two stage least squares.
[a] *Robust-unchanged* refers to the situation where the estimates of one or more parameters are identical for two different model structures. See Bollen et al. (2018) for further discussion of the meaning of robust to structural misspecifications.

ROBUSTNESS

2. Structural misspecification robustness

- omitted paths
- omitted variables
- wrong number of dimensions

Bollen (2001): Suppose that for the j^{th} equation in the correctly specified model, the model implied IVs are in a matrix $\mathbf{V}_j$. The 2SLS estimator of the coefficients is robust for any misspecification in other equations under two conditions:

1. The equation being estimated is correctly specified
2. The misspecifications in the other equations do not alter the variables in $\mathbf{V}_j$

ROBUSTNESS

2. Structural misspecification robustness

- omitted paths
- omitted variables
- wrong number of dimensions

After demonstrating the use of lavaan's simulateData() function we'll explore the consequences of structural misspecifications on the MIIV-2SLS estimates.

First, we'll walk through one full example together.

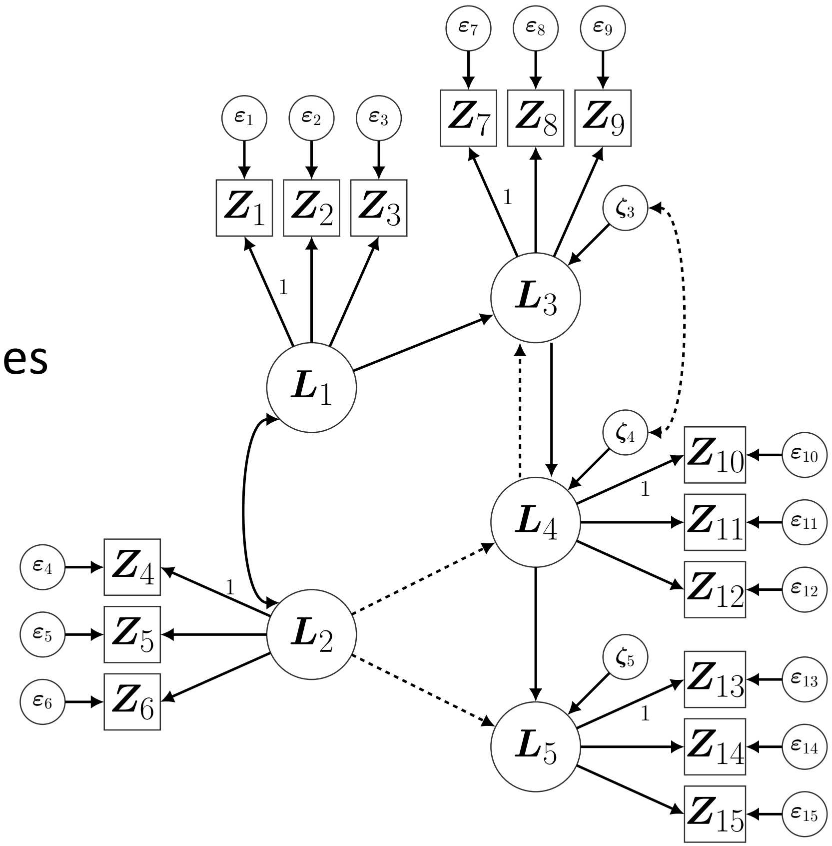
ROBUSTNESS

Measurement Model Robustness to Latent Variable Model Misspecification

True Model :
- Solid and dashed lines

Misspecifications
- Omitted paths, omitted latent variables
- Omitted covariances of errors

Simulation
- Simulated data to true model
- 1,000 observations
- 3 indicators per latent variable
- Estimate measurement model 2x
  1. Latent variable model is correct (solid and dashed lines)
  2. Latent variable model is incorrect(omitting dashed lines)



ROBUSTNESS

Use lavaan to simulate data

- `simulateData` function from lavaan.

Need to load lavaan

```
library("lavaan")
```

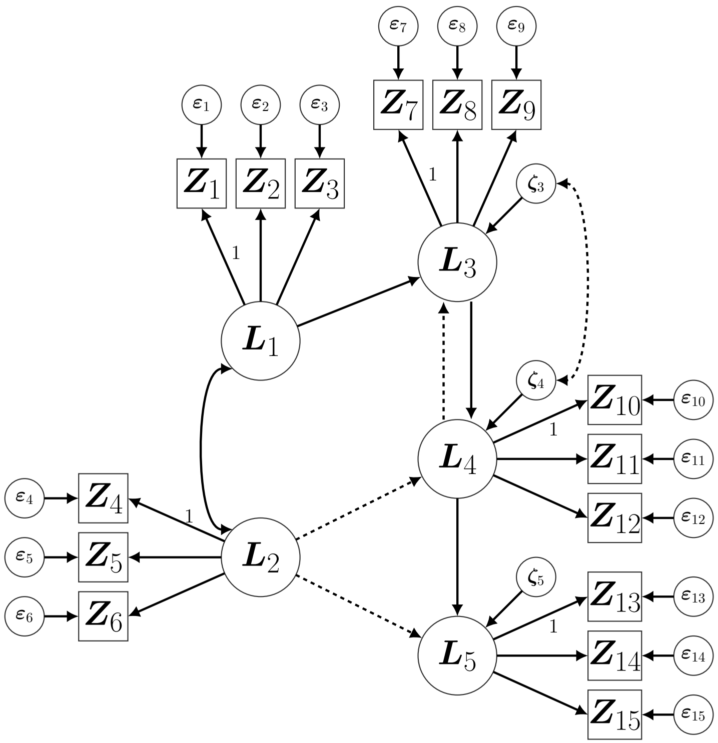
Specify the model from the previous slide

- Include population parameters
- Name this model: `model.sim.1`.

ROBUSTNESS

Here we specify the data generating model.

Note: Var(error) & Var(exogenous) =1 by default.




```
library("lavaan")

model.sim.1 <- '
  L1  =~ 1*Z1  + 1*Z2  + 1*Z3
  L2  =~ 1*Z4  + 1*Z5  + 1*Z6
  L3  =~ 1*Z7  + 1*Z8  + 1*Z9
  L4  =~ 1*Z10 + 1*Z11 + 1*Z12
  L5  =~ 1*Z13 + 1*Z14 + 1*Z15

  L3 ~  .3*L1 + .5*L4
  L4 ~  .5*L2 + .8*L3
  L5 ~  .3*L2 + .5*L4

  L3 ~~ .6*L4'
```

ROBUSTNESS

Now use lavaan's simulateData function

- Set the model argument to `model`.sim.1
- Fix the number of observations to 1,000
- Choose random.seed of 123
  - Allows us to replicate our results exactly
- Save our dataset as data.sim.1

```
data.sim.1 <- simulateData(model = model.sim.1,
                           sample.nobs = 1000,
                           seed = 123)
```

ROBUSTNESS

Below we recap the commands used to simulate our data set.

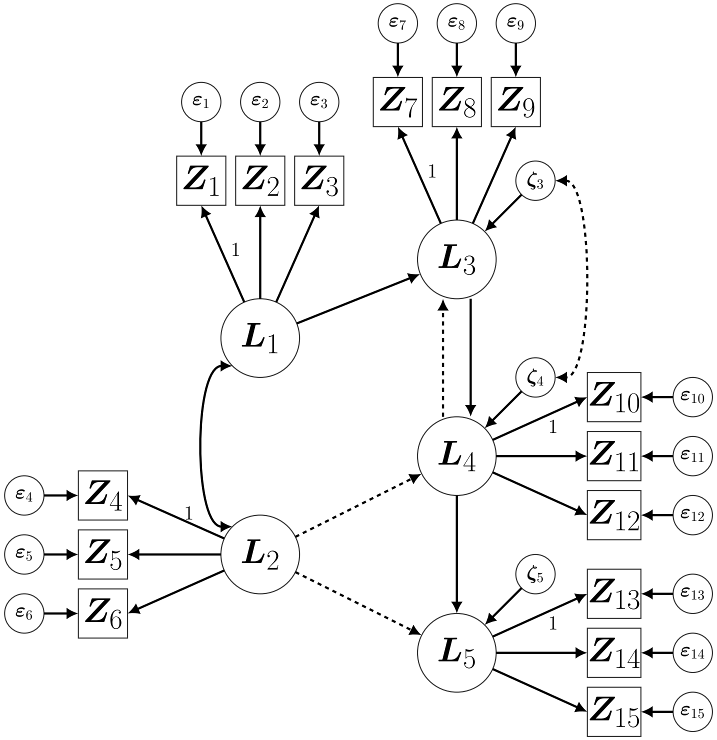
```
library("lavaan")

model.sim.1 <- '
  L1  =~ 1*Z1  + 1*Z2  + 1*Z3
  L2  =~ 1*Z4  + 1*Z5  + 1*Z6
  L3  =~ 1*Z7  + 1*Z8  + 1*Z9
  L4  =~ 1*Z10 + 1*Z11 + 1*Z12
  L5  =~ 1*Z13 + 1*Z14 + 1*Z15

  L3 ~  .3*L1 + .5*L4
  L4 ~  .5*L2 + .8*L3
  L5 ~  .3*L2 + .5*L4

  L3 ~~ .6*L4
'
data.sim.1 <- simulateData(model = model.sim.1,
                           sample.nobs = 1000,
                           seed = 123)
```

Lastly, we need to define two estimating models:

1. `model.correct.1`
2. `model.misspecified.1`

`model.correct.1` corresponds to the correct model (both solid and dashed lines).

```
model.correct.1 <- '

  L1  =~ Z1  + Z2  + Z3
  L2  =~ Z4  + Z5  + Z6
  L3  =~ Z7  + Z8  + Z9
  L4  =~ Z10 + Z11 + Z12
  L5  =~ Z13 + Z14 + Z15

  L3 ~  L1 + L4
  L4 ~  L2 + L3
  L5 ~  L2 + L4

  L3 ~~ L4
'
```

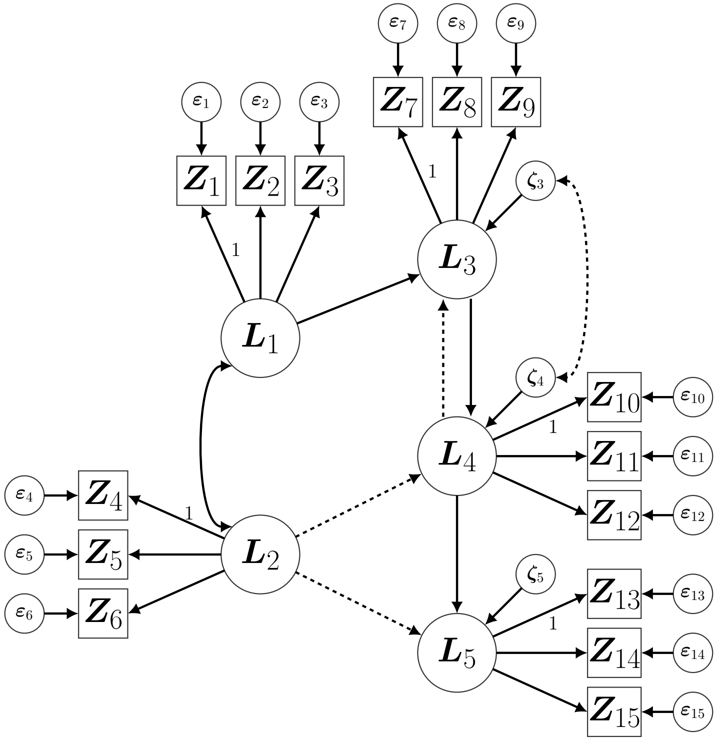


Lastly, we need to define two estimating models:

1. `model.correct.1`
2. `model.misspecified.1`

`model.misspecified.1` corresponds to the incorrect model (solid lines only).

```
model.misspecified.1 <- '

  L1  =~ Z1  + Z2  + Z3
  L2  =~ Z4  + Z5  + Z6
  L3  =~ Z7  + Z8  + Z9
  L4  =~ Z10 + Z11 + Z12
  L5  =~ Z13 + Z14 + Z15

  L3 ~  L1
  L4 ~  L3
  L5 ~  L4
'
```



ROBUSTNESS

At this point we should have defined both the correct and misspecified models.

```
model.correct.1 <- '

  L1  =~ Z1  + Z2  + Z3
  L2  =~ Z4  + Z5  + Z6
  L3  =~ Z7  + Z8  + Z9
  L4  =~ Z10 + Z11 + Z12
  L5  =~ Z13 + Z14 + Z15

  L3 ~  L1 + L4
  L4 ~  L2 + L3
  L5 ~  L2 + L4

  L3 ~~ L4
'
```

```
model.misspecified.1 <- '

  L1  =~ Z1  + Z2  + Z3
  L2  =~ Z4  + Z5  + Z6
  L3  =~ Z7  + Z8  + Z9
  L4  =~ Z10 + Z11 + Z12
  L5  =~ Z13 + Z14 + Z15

  L3 ~  L1
  L4 ~  L3
  L5 ~  L4
'
```

ROBUSTNESS

Estimate correct & incorrect models in MIIVsem.

The `estimatesTable()` function

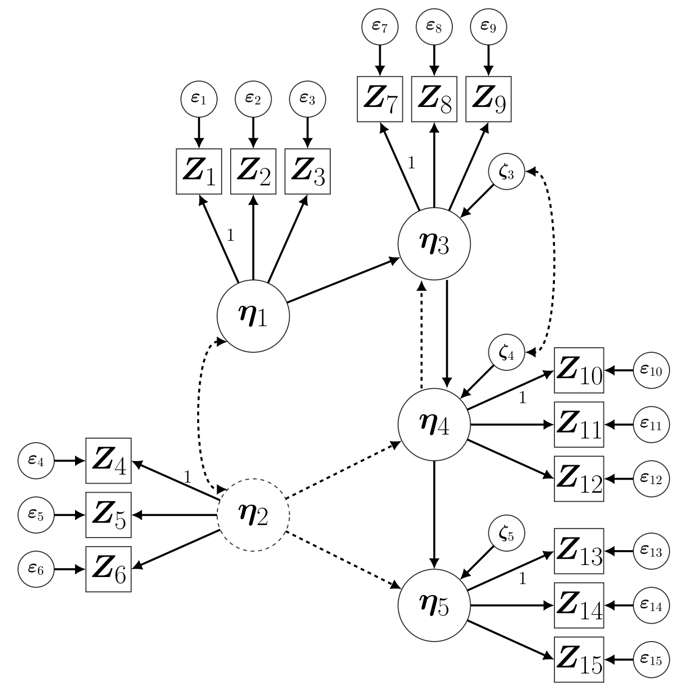
- Convenient way to store, view and manipulate the estimated parameters returned by `miive()`
- Save the fitted correct and misspecified models as `fit.cor.1`, and `fit.mis.1`, respectively.

```
fit.cor.1 <- miive(model.correct.1, data.sim.1)
fit.mis.1 <- miive(model.misspecified.1, data.sim.1)

estimatesTable(fit.cor.1)
estimatesTable(fit.mis.1)
```

Simulation 1 Results

Measurement Model Robustness to Latent Variable Model Misspecification

```
              est.cor.1 est.mis.1
1  L1 =~  Z1 1.0000000 1.0000000
2  L1 =~  Z2 1.0873919 1.0873919
3  L1 =~  Z3 1.0302098 1.0302098
4  L2 =~  Z4 1.0000000 1.0000000
5  L2 =~  Z5 1.0076301 1.0076301
6  L2 =~  Z6 0.9691372 0.9691372
7  L3 =~  Z7 1.0000000 1.0000000
8  L3 =~  Z8 0.9933688 0.9933688
9  L3 =~  Z9 0.9791256 0.9791256
10 L4 =~ Z10 1.0000000 1.0000000
11 L4 =~ Z11 1.0159817 1.0159817
12 L4 =~ Z12 1.0228551 1.0228551
13 L5 =~ Z13 1.0000000 1.0000000
14 L5 =~ Z14 1.0019276 1.0019276
15 L5 =~ Z15 0.9667204 0.9667204
```



EXERCISE: Simulation 2

Previously we examined the impact of a misspecified latent variable model on the measurement model estimates. We now consider another situation, the impact of a misspecified measurement model on the latent variable model estimates.

For this exercise you will:

1. Use path diagram of next slide and lavaan's simulateData function to simulate 1,000 observations. In the data generating model set the regression coefficients and factor loadings to a value of 1, and the uniqueness covariances to a value of 0.3.
2. Fit the correct and misspecified models using MIIVsem.
3. Compare the estimates.

EXERCISE: Simulation 2

Latent Variable Model Robustness to Measurement Model Misspecification

True Model :

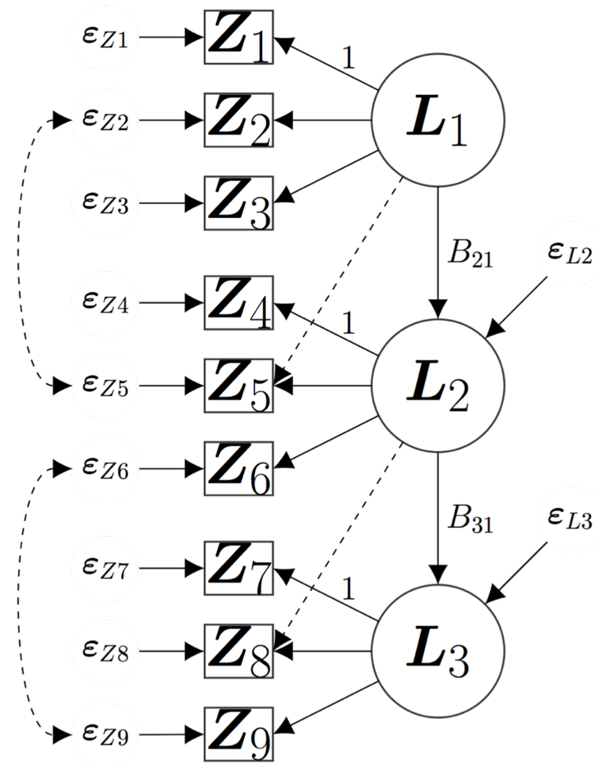
- Solid and dashed lines

Misspecifications

- Omitted correlated errors of indicators
- Omitted cross-loadings

Simulation

- Simulated data to true model
- 1,000 observations
- 3 indicators per latent variable
- Estimate latent variable model 2x
  1. Measurement model is correct (solid and dashed lines)
  2. Measurement model is incorrect (omitting dashed lines)



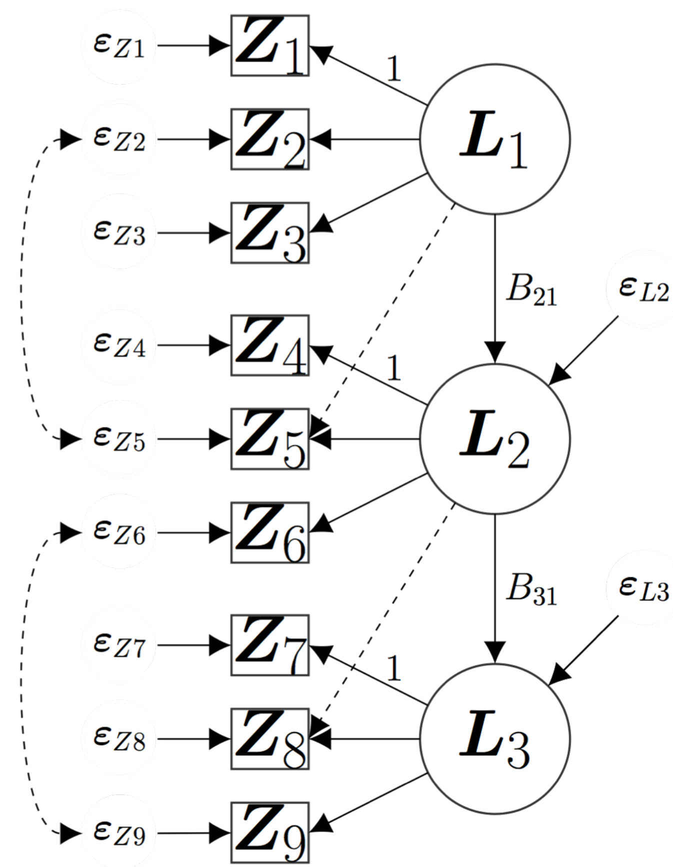
SOLUTION: Simulation 2

To simulate data for our example we will use the `simulateData` function from lavaan.

```
library("lavaan")

model.sim.2 <- '

  L1  =~ 1*Z1  + 1*Z2  + 1*Z3 + 1*Z5
  L2  =~ 1*Z4  + 1*Z5  + 1*Z6 + 1*Z8
  L3  =~ 1*Z7  + 1*Z8  + 1*Z9

  L2  ~  1*L1
  L3  ~  1*L2

  Z2 ~~ .3*Z5
  Z6 ~~ .3*Z9'

data.sim.2 <- simulateData(model = model.sim.2,
                           sample.nobs = 1000,
                           seed = 123)
```

EXERCISE: Simulation 2

- Now that we have simulated the data we'll have to specify the models.
- The correct model includes the solid and dashed lines. The misspecified model contains the dashed lines only.



SOLUTION: Simulation 2

Now, we can specify the correct (solid and dashed lines) and misspecified (solid lines only) models.

```
model.correct.2 <- '

  L1  =~ Z1  + Z2  + Z3 + Z5
  L2  =~ Z4  + Z5  + Z6 + Z8
  L3  =~ Z7  + Z8  + Z9

  L2  ~  L1
  L3  ~  L2

  Z2  ~~ Z5
  Z6  ~~ Z9
'
```

```
model.misspecified.2 <- '

  L1  =~ Z1  + Z2  + Z3
  L2  =~ Z4  + Z5  + Z6
  L3  =~ Z7  + Z8  + Z9

  L2 ~  L1
  L3 ~  L2
'
```

SOLUTION: Simulation 2

Correctly Specified Model (Unadjusted Sargan Test)

STRUCTURAL COEFFICIENTS:

	Estimate	Std.Err	z-value	P(>|z|)	Sargan	df	P(Chi)
L1 =~							
Z1	1.000						
Z2	0.927	0.046	20.026	0.000	13.909	5	0.016
Z3	0.928	0.044	20.871	0.000	4.656	6	0.589
Z5	0.766	0.120	6.402	0.000	3.959	3	0.266
L2 =~							
Z4	1.000						
Z5	1.117	0.084	13.339	0.000	3.959	3	0.266
Z6	0.969	0.034	28.863	0.000	9.882	5	0.079
Z8	0.892	0.092	9.660	0.000	5.790	4	0.215
L3 =~							
Z7	1.000						
Z8	1.125	0.083	13.533	0.000	5.790	4	0.215
Z9	1.028	0.031	33.274	0.000	6.722	5	0.242
L2 ~							
L1	0.933	0.055	16.861	0.000	0.059	1	0.808
L3 ~							
L2	0.922	0.041	22.242	0.000	8.924	4	0.063

SOLUTION: Simulation 2

Correctly Specified Model (Adjusted Sargan Test)

STRUCTURAL COEFFICIENTS:

	Estimate	Std.Err	z-value	P(>|z|)	Sargan	df	P(Chi)
L1 =~							
Z1	1.000						
Z2	0.927	0.046	20.026	0.000	13.909	5	0.130
Z3	0.928	0.044	20.871	0.000	4.656	6	1.000
Z5	0.766	0.120	6.402	0.000	3.959	3	1.000
L2 =~							
Z4	1.000						
Z5	1.117	0.084	13.339	0.000	3.959	3	1.000
Z6	0.969	0.034	28.863	0.000	9.882	5	0.472
Z8	0.892	0.092	9.660	0.000	5.790	4	1.000
L3 =~							
Z7	1.000						
Z8	1.125	0.083	13.533	0.000	5.790	4	1.000
Z9	1.028	0.031	33.274	0.000	6.722	5	1.000
L2 ~							
L1	0.933	0.055	16.861	0.000	0.059	1	1.000
L3 ~							
L2	0.922	0.041	22.242	0.000	8.924	4	0.441

SOLUTION: Simulation 2

Misspecified Model (Unadjusted Sargan Test)

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
  L1 =~
    Z1                1.000
    Z2                0.981    0.044   22.149    0.000   22.579    6    0.001
    Z3                0.928    0.044   20.871    0.000    4.656    6    0.589
  L2 =~
    Z4                1.000
    Z5                1.602    0.047   33.958    0.000  104.349    6    0.000
    Z6                0.973    0.034   28.952    0.000   35.800    6    0.000
  L3 =~
    Z7                1.000
    Z8                1.849    0.050   37.256    0.000   64.561    6    0.000
    Z9                1.026    0.031   33.240    0.000   44.087    6    0.000

  L2 ~
    L1                0.933    0.055   16.861    0.000    0.059    1    0.808
  L3 ~
    L2                0.922    0.041   22.242    0.000    8.924    4    0.063
```

SOLUTION: Simulation 2

Misspecified Model (Adjusted Sargan Test)

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
  L1 =~
    Z1                1.000
    Z2                0.981    0.044   22.149    0.000   22.579    6    0.004
    Z3                0.928    0.044   20.871    0.000    4.656    6    1.000
  L2 =~
    Z4                1.000
    Z5                1.602    0.047   33.958    0.000  104.349    6    0.000
    Z6                0.973    0.034   28.952    0.000   35.800    6    0.000
  L3 =~
    Z7                1.000
    Z8                1.849    0.050   37.256    0.000   64.561    6    0.000
    Z9                1.026    0.031   33.240    0.000   44.087    6    0.000

  L2 ~
    L1                0.933    0.055   16.861    0.000    0.059    1    1.000
  L3 ~
    L2                0.922    0.041   22.242    0.000    8.924    4    0.189
```

SOLUTION: Simulation 2

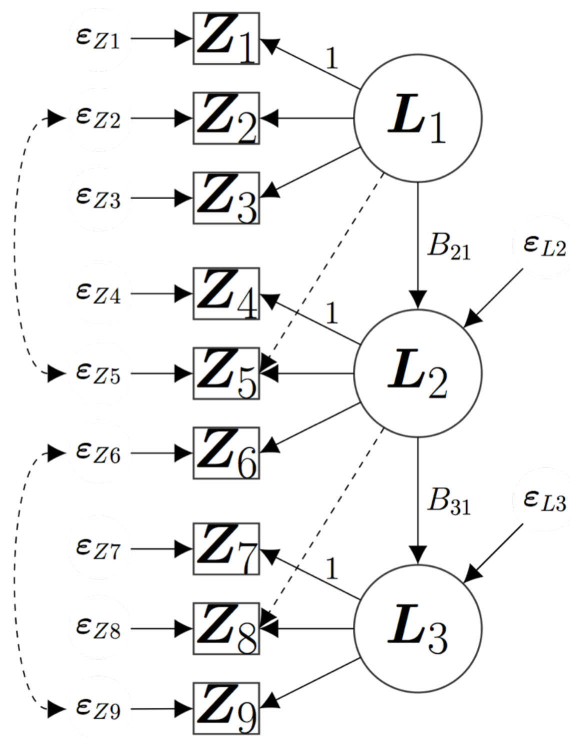
Below is code for combining estimates from more than one fitted models for easy comparison .

```
# Save estimated models
fit.cor.2 <- miive(model.correct.2, data.sim.2)
fit.mis.2 <- miive(model.misspecified.2, data.sim.2)

# Save parameter tables
est.cor.2 <- estimatesTable(fit.cor.2)
est.mis.2 <- estimatesTable(fit.mis.2)

# Merge tables
compare.2 <- merge(
   est.cor.2[est.cor.2$op == "~", c("lhs","op","rhs","est")],
   est.mis.2[est.mis.2$op == "~", c("lhs","op","rhs","est")],
   by = c("lhs","op","rhs")
)

colnames(compare.2) <- c("","","","est.cor.2","est.mis.2")
compare.2
```

SOLUTION: Simulation 2

Measurement Model Robustness to Latent Variable Model Misspecification

```
##              est.cor.2 est.mis.2
## 1 L2 ~ L1 0.9332420 0.9332420
## 2 L3 ~ L2 0.9222349 0.9222349
```



EXERCISE: Simulation 3

We now consider a final simulated data example. Here we make a different measurement model misspecification and ask you to answer the following three questions:

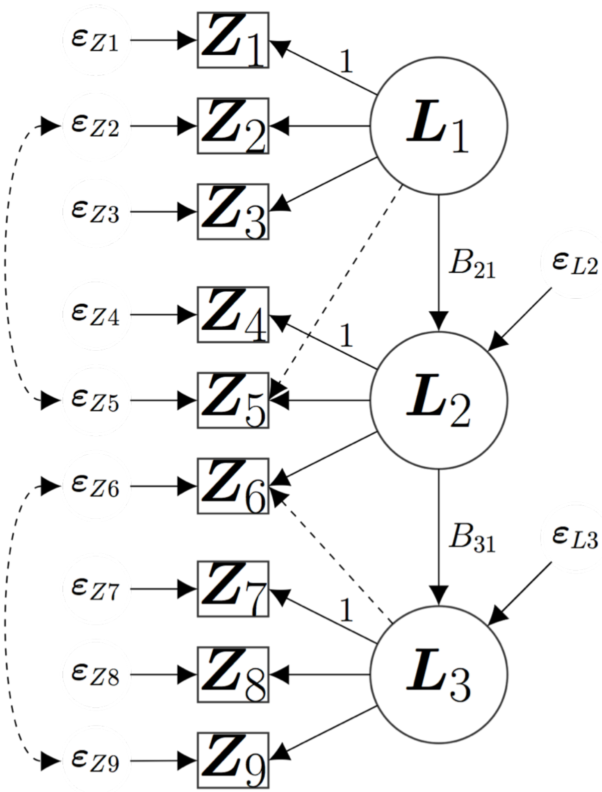
1. Do the robustness properties discussed earlier hold here?
2. If not, what is the reason for the change?
3. Is there any way of detecting this misspecification in practice?

Path diagram and questions presented on next slide.

EXERCISE: Simulation 3

Robustness Exercise

- Simulated 1,000 observations according to true model (solid and dashed lines).
- Estimate two models:
  1. Model 1: True model
  2. Model 2: Misspecified model where the dashed paths are omitted.



Questions

1. Do the robustness properties discussed earlier hold here?
2. If not, what is the reason for the change?
3. Is there any way of detecting this misspecification in practice?

SOLUTION: Simulation 3

Correctly Specified Model (Adjusted Sargan)

STRUCTURAL COEFFICIENTS:

	Estimate	Std.Err	z-value	P(>|z|)	Sargan	df	P(Chi)
L1 =~							
Z1	1.000						
Z2	0.954	0.047	20.215	0.000	7.952	5	0.953
Z3	0.921	0.044	20.718	0.000	3.205	6	1.000
Z5	0.784	0.138	5.673	0.000	3.211	3	1.000
L2 =~							
Z4	1.000						
Z5	1.094	0.094	11.599	0.000	3.211	3	1.000
Z6	0.930	0.091	10.179	0.000	2.893	3	1.000
L3 =~							
Z7	1.000						
Z6	1.001	0.078	12.757	0.000	2.893	3	1.000
Z8	0.960	0.027	35.206	0.000	8.561	6	0.999
Z9	0.978	0.029	33.990	0.000	10.685	5	0.464
L2 ~							
L1	1.006	0.058	17.382	0.000	0.434	1	1.000
L3 ~							
L2	0.930	0.042	22.001	0.000	7.257	3	0.464

SOLUTION: Simulation 3

Misspecified Model (Adjusted Sargan)

STRUCTURAL COEFFICIENTS:

	Estimate	Std.Err	z-value	P(>|z|)	Sargan	df	P(Chi)
L1 =~							
Z1	1.000						
Z2	1.015	0.045	22.721	0.000	17.476	6	0.031
Z3	0.921	0.044	20.718	0.000	3.205	6	1.000
L2 =~							
Z4	1.000						
Z5	1.613	0.049	32.645	0.000	84.966	6	0.000
Z6	2.034	0.061	33.073	0.000	113.670	6	0.000
L3 =~							
Z7	1.000						
Z8	0.960	0.027	35.206	0.000	8.561	6	0.599
Z9	1.012	0.028	36.543	0.000	22.987	6	0.004
L2 ~							
L1	1.006	0.058	17.382	0.000	0.434	1	1.000
L3 ~							
L2	1.073	0.042	25.418	0.000	109.497	4	0.000

ROBUSTNESS

2. Structural misspecification robustness
   - MIIV-2SLS is robust because the MIIVs are the same for all models
   - MIIV-2SLS depends on identification of equation, not identification of whole model
   - MIIV-2SLS is NOT robust to all structural misspecifications
     - E.g., the measurement model estimates are not robust to the different models illustrated.

DIMENSIONALITY

In the next set of slides we show how MIIV-2SLS can be used to investigate dimensionality.

In doing so we demonstrate two features in MIIVsem:

1. How to add constraints to parameters in the lavaan style model syntax.
2. How to obtain Wald tests of these restrictions.

The data are included in MIIVsem (`bollen1996`) and come from a survey conducted in rural clusters of Tanzania to collect information on the perceived accessibility of a specific family planning facility that serviced each cluster (Bollen, 1996).

DIMENSIONALITY

Background Information:

Six informants were chosen from each cluster: 3 female and 3 male. New informants were chosen for each cluster. Each informant was independently asked to rate the accessibility of the facility. More specifically the women informants were asked to rate how women of childbearing age perceived the accessibility of the clinic and the men informants were asked to rate how accessible men perceived the clinic to be.

The female informants' ratings are `access1-3` and the male informants' ratings are `access4-6`.

DIMENSIONALITY

Model 1

We will begin by fitting a confirmatory factor analysis model with a single accessibility latent variable for male and female information.

Since the informants change from cluster to cluster, the ordering of the informants is arbitrary, and we have no reason to believe the factor loadings would differ in a systematic way.

In Model 1 we will constrain all the factor loadings to equality.

DIMENSIONALITY

Equality Constraints and Parameter Restrictions

Identical to lavaan, labels can be used to specify equality constraints on parameters in the model syntax. Labels are prepended to the variable name using the `*` operator. For numeric constrains one can specify a number instead.

```
model.1 <- '

  accessibility =~  1*access1 + 1*access2 + 1*access3 +
                    1*access4 + 1*access5 + 1*access6

'
```

By constraining each loading to 1 we constrain all the loadings to equality.

DIMENSIONALITY

MIIVsem test statistics for constraints

- Large-sample Wald test of constraints imposed on coefficient matrix

F and χ^2 distributed test statistics asymptotically equivalent

- Performance may differ in small samples
- See Greene (2003, pp. 346-347) for details

Conduct Wald test by saving the miive object and using the `summary()` method request `restrict.tests=TRUE`

```
fit <- miive(model.1, bollen1996, sarg.adjust = "holm")

summary(fit, restrict.tests = TRUE)
```

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
  accessibility =~
    access1           1.000
    access2           1.000                                3.800    3    0.568
    access3           1.000                                1.107    3    0.775
    access4           1.000                               71.309    3    0.000
    access5           1.000                               82.893    3    0.000
    access6           1.000                               72.377    3    0.000

MIIV-2SLS LINEAR HYPOTHESIS TESTS:

   access2_access1 = 1
   access3_access1 = 1
   access4_access1 = 1
   access5_access1 = 1
   access6_access1 = 1

   Wald Test (Chi^2):                             18.1306
   Degrees of freedom:                                  5
   Pr(>Chi^2) :                                    0.0028

   Wald Test (F):                                  3.6261
   Degrees of freedom:                               5, 5
   Pr(>F) :                                        0.0029
```

EXERCISE: Dimensionality 1

Model 2

The informants change from cluster to cluster but we have evidence the factor loadings for the female informants and those for the male informants are not equal. Fit a model where females have the same loading, and males loadings are also constrained equal. Note males loadings can differ from females. What has changed across the two model?

```
model.2 <- ‘
  accessibility =~  1*access1 + 1*access2 + 1*access3 +
                    b*access4 + b*access5 + b*access6
‘
miive(model.2, bollen1996, sarg.adjust = "holm")
```

SOLUTION: Dimensionality 1

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
  accessibility =~
    access1           1.000
    access2           1.000                                3.800    3    0.568
    access3           1.000                                1.107    3    0.775
    access4           0.727    0.065   11.249    0.000    82.304    3    0.000
    access5           0.727    0.065   11.249    0.000    95.313    3    0.000
    access6           0.727    0.065   11.249    0.000    82.175    3    0.000


MIIV-2SLS LINEAR HYPOTHESIS TESTS:

   access2_access1 = 1
   access3_access1 = 1
   access4_access1 - access5_access1 = 0
   access4_access1 - access6_access1 = 0

   Wald Test (Chi^2):                         0.3103
   Degrees of freedom:                             4
   Pr(>Chi^2) :                               0.9891

   Wald Test (F):                             0.0776
   Degrees of freedom:                          4, 4
   Pr(>F) :                                   0.9891
```

EXERCISE: Dimensionality 2

Model 3

If women and men differ in their view of accessibility, then a two factor model could be more appropriate. Create a model that has Female Accessibility as the first factor and Male Accessibility as the second factor. Allow the two factors to correlate.

Scale the Female Accessibility latent variable to access1 and the Male Accessibility latent variable to access4. What happens to the overidentification tests?

SOLUTION: Dimensionality 2

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
  Female =~
    access1           1.000
    access2           1.041    0.104   10.039    0.000    3.499    3    0.963
    access3           1.033    0.103   10.034    0.000    0.970    3    0.963
  Male =~
    access4           1.000
    access5           1.026    0.073   14.039    0.000    6.623    3    0.340
    access6           1.005    0.076   13.285    0.000    2.581    3    0.963
```

CATEGORICAL ENDOGENOUS VARIABLES

Polychoric Instrumental Variable (PIV) Estimator

- Access indicators are ordinal
  - what happens if take account of ordinal nature?
- Polychoric correlations
  - assumes continuous underlying variable
  - ordinal variables are collapsed version
  - estimates of correlation between underlying variables
  - polychoric correlations then analyzed
- MIIVsem permits endogenous ordinal variables analyzed with polychoric correlations
  - Polychoric Instrumental Variable (PIV) estimator
  - See Bollen & Maydeu-Olivares (2007), Fisher & Bollen (2020) for details

CATEGORICAL ENDOGENOUS VARIABLES

Following the convention used in lavaan, we use the ordered argument to indicate which variables in the model syntax are categorical. This will be demonstrated in the following example.

In our previous Accessibility example, responses were recorded on a 1-5 scale. We will reestimate our final two-factor model to demonstrate the PIV estimator.

CATEGORICAL ENDOGENOUS VARIABLES

Specify the model:

```
model <- ‘
   femaleAccess  =~ access1 + access2 + access3
   maleAccess    =~ access4 + access5 + access6
‘
```

Declare the ordered categorical variables:

```
ordered <- c("access1", "access2","access3",
             "access4", "access5","access6")
```

Fit the model:

```
miive(model, bollen1996, ordered = ordered)
```

```
MIIVsem (0.5.2) results

Number of observations                                                        220
Number of equations                                                             4
Estimator                                                          MIIV-2SLS (PIV)
Standard Errors                                                          standard
Missing                                                                  listwise


Parameter Estimates:


STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
 Female =~
    access1           1.000
    access2           1.039    0.092   11.313    0.000
    access3           0.983    0.078   12.680    0.000
  Male =~
    access4           1.000
    access5           1.026    0.086   11.976    0.000
    access6           0.967    0.072   13.355    0.000
```

Note: In the header, the estimator is now listed as MIIV-2SLS (PIV)

Reisenzein (1986)

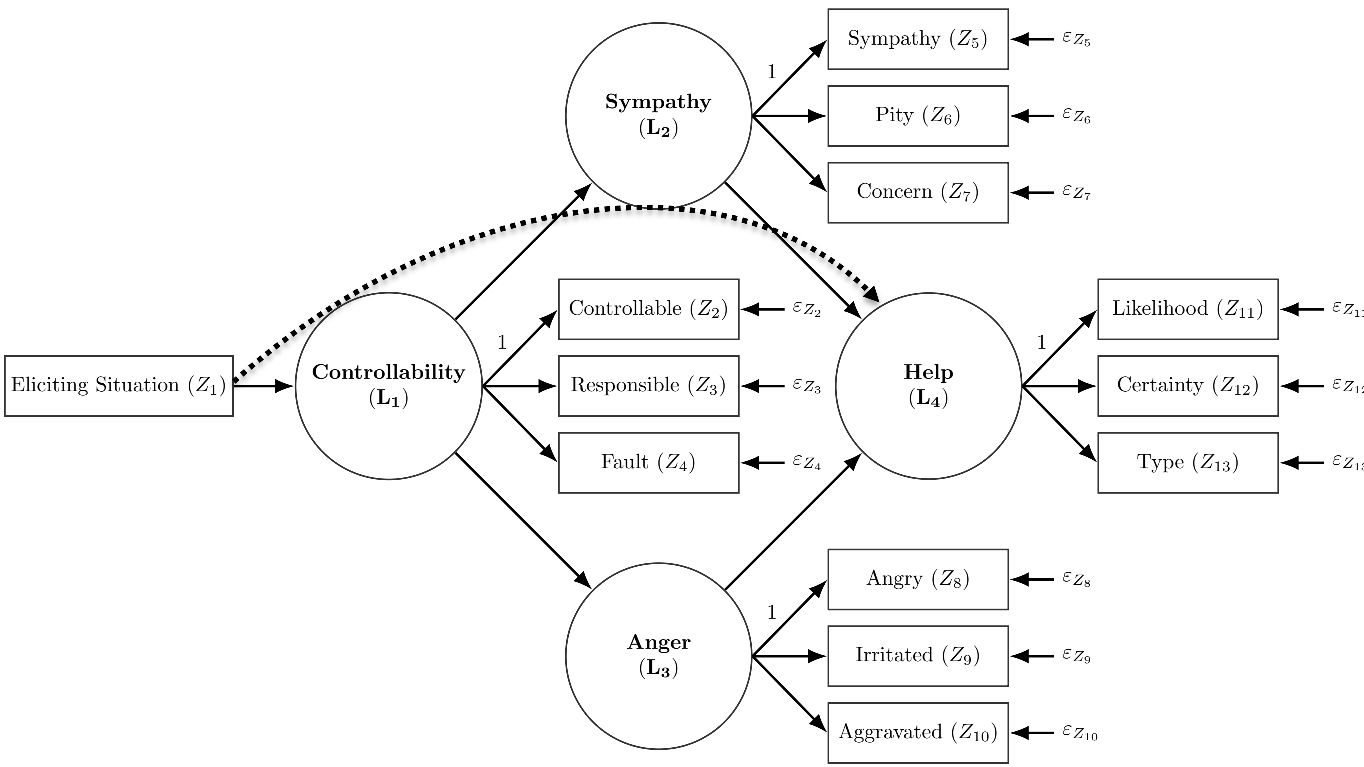
The final example uses data from an experiment conducted by Reisenzein (1986).

Professor Reisenzein has kindly allowed us to include his data in MIIVsem.

The data is saved as `reisenzein1986`.

Reisenzein (1986)

Finally, using the path diagram below fit the models described by Reisenzein (1986). The first model contains solid lines only, the second contains both the solid and dashed lines.



SOLUTION: Reisenzein (1986)

Model 1 Syntax: Figure 2 without dashed line

```
model.reisenzein.1 <- '
  L2 =~ Z5 + Z6 + Z7
  L1 =~ Z2 + Z3 + Z4
  L3 =~ Z8 + Z9 + Z10
  L4 =~ Z11 + Z12 + Z13

  L4 ~ L2 + L3
  L2 ~ L1
  L3 ~ L1
  L1 ~ Z1
'
```

SOLUTION: Reisenzein (1986)

Fit Model 1 using the MIIV-2SLS estimator:

```
miive(model.reisenzein.1, reisenzein1986, sarg.adjust = “holm”)
```

SOLUTION: Reisenzein (1986)

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
  L1 =~
    Z2                1.000
    Z3                1.047    0.080   13.041    0.000   12.085   10    1.000
    Z4                1.152    0.087   13.218    0.000   11.550   10    1.000
  L2 =~
    Z5                1.000
    Z6                0.724    0.068   10.594    0.000   10.419   10    1.000
    Z7                0.721    0.061   11.779    0.000   26.677   10    0.032
  L3 =~
    Z8                1.000
    Z9                0.895    0.075   12.009    0.000    5.281   10    1.000
    Z10               0.885    0.071   12.474    0.000    9.269   10    1.000
  L4 =~
    Z11               1.000
    Z12               1.098    0.059   18.746    0.000    7.632   10    1.000
    Z13               0.428    0.034   12.686    0.000   11.850   10    1.000

  L1 ~
    Z1                3.828    0.315   12.142    0.000
  L2 ~
    L1               -0.716    0.092   -7.780    0.000   10.436    5    0.510
  L3 ~
    L1                0.640    0.078    8.245    0.000   14.644    5    0.120
  L4 ~
    L2                0.430    0.082    5.270    0.000   16.016    6    0.123
    L3               -0.405    0.094   -4.320    0.000
```

SOLUTION: Reisenzein (1986)

Model 2 Syntax: Figure 2 with dashed line

```
model.reisenzein.2 <- '
  L2 =~ Z5 + Z6 + Z7
  L1 =~ Z2 + Z3 + Z4
  L3 =~ Z8 + Z9 + Z10
  L4 =~ Z11 + Z12 + Z13

  L4 ~ L2 + L3 + Z1
  L2 ~ L1
  L3 ~ L1
  L1 ~ Z1
 '
```

SOLUTION: Reisenzein (1986)

Fit Model 2 using the MIIV-2SLS estimator:

```
miive(model.reisenzein.2, reisenzein1986, sarg.adjust = “holm”)
```

SOLUTION: Reisenzein (1986)

STRUCTURAL COEFFICIENTS:

	Estimate	Std.Err	z-value	P(>|z|)	Sargan	df	P(Chi)
L1 =~							
Z2	1.000						
Z3	1.047	0.080	13.041	0.000	12.085	10	1.000
Z4	1.152	0.087	13.218	0.000	11.550	10	1.000
L2 =~							
Z5	1.000						
Z6	0.724	0.068	10.594	0.000	10.419	10	1.000
Z7	0.721	0.061	11.779	0.000	26.677	10	0.032
L3 =~							
Z8	1.000						
Z9	0.895	0.075	12.009	0.000	5.281	10	1.000
Z10	0.885	0.071	12.474	0.000	9.269	10	1.000
L4 =~							
Z11	1.000						
Z12	1.098	0.059	18.746	0.000	7.632	10	1.000
Z13	0.428	0.034	12.686	0.000	11.850	10	1.000
L1 ~							
Z1	3.828	0.315	12.142	0.000			
L2 ~							
L1	-0.716	0.092	-7.780	0.000	10.436	5	0.510
L3 ~							
L1	0.640	0.078	8.245	0.000	14.644	5	0.108
L4 ~							
Z1	-0.712	0.462	-1.540	0.124	15.040	5	0.102
L2	0.376	0.086	4.370	0.000			
L3	-0.308	0.110	-2.812	0.005			

Questions from Reisenzein (1986)

Variable name	Construct	Question or description
Z1	None	Eliciting situation
Z2	Controllability	How controllable, do you think, is the cause of the person's present condition? (1 = *not at all under personal control*, 9 = *completely under personal control*).
Z3		How responsible, do you think, is that person for his present condition? (1 = *not at all responsible*, 9 = *very much responsible*).
Z4		I would think that it was the person's own fault that he is in the present situation. (1 = *no. not at all*. 9 = *yes, absolutely so*).
Z5	Sympathy	How much sympathy would you feel for that person? (1 = *none at all*, 9 = *very much*).
Z6		I would feel pity for this person. (1 = *none at all*, 9 = *very much*).
Z7		How much concern would you feel for this person? (1 = *none at all*, 9 = *very much*).
Z8	Anger	How angry would you feel at that person? (1 = *not at all*, 9 = *very much*).
Z9		How irritated would you feel by that person? (1 = *not at all*, 9 = *very much*).
Z10		I would feel aggravated by that person. (1 = *not at all*, 9 = *very much so*).
Z11	Help	How likely is it that you would help that person? (1 = *definitely would not help*, 9 = *definitely would help*).
Z12		How certain would you feel that you would help the person? (1 = *not at all certain*, 9 = *absolutely certain*).
Z13		Which of the following actions would you most likely engage in? 1 = *not help at all, try to stay uninvolved*; 2 = *try to alert other bystanders, but stay uninvolved myself*; 3 = *try to inform the conductor or another official in charge*; 4 = *go over and help the person to a seat*; 5 = *help in any way that might be necessary, including if necessary first aid and/or accompanying the person to a hospital*.

[a] Variable symbols correspond to path diagram and equations in our article. Descriptions of variables taken from Reisenzein (1986).

Table 3 from Reisenzein (1986)

Model	χ^2 (N = 138)	df	p
Null model	1,463.94	78	<.001
Model 1	86.58	61	<.020
Model 2 (with β_{HC} added)	85.95	60	<.016
Model 3 (with a correlation between the error terms of sympathy and anger added)	85.44	60	<.018
Model 4 (with both β_{HC} and the correlation between the errors of sympathy and anger added)	84.97	59	<.016
Model 5 (with β_{HE} added)	82.25	60	<.029
Model 5′ (= 5 with 2 correlated error terms)	67.28	58	≅.189
Model comparisons			
1 versus 2	0.63	1	$.50 > p > .25$
1 versus 3	1.14	1	$.50 > p > .25$
1 versus 4	1.61	2	$.50 > p > .25$
1 versus 5	4.33	1	<.05
1 versus null	1,377.36	17	<.001
5 versus null	1.69	18	<.001
5 versus 5′	14.97	2	<.001

SOLUTION: Reisenzein (1986)

Cross-loading Model 1 Syntax:

```
model.cross.1 <- '
  L2 =~ Z5 + Z6 + Z7
  L1 =~ Z2 + Z3 + Z4
  L3 =~ Z8 + Z9 + Z10
  L4 =~ Z11 + Z12 + Z13 + Z7

  L4 ~ L2 + L3
  L2 ~ L1
  L3 ~ L1
  L1 ~ Z1
 '
```

SOLUTION: Reisenzein (1986)

Fit Cross–Loading Model 1:

```
miive(model.cross.1, reisenzein1986, , sarg.adjust = "holm")
```

SOLUTION: Reisenzein (1986)

STRUCTURAL COEFFICIENTS:

	Estimate	Std.Err	z-value	P(>|z|)	Sargan	df	P(Chi)
L1 =~							
Z2	1.000						
Z3	1.047	0.080	13.041	0.000	12.085	10	1.000
Z4	1.152	0.087	13.218	0.000	11.550	10	1.000
L2 =~							
Z5	1.000						
Z6	0.724	0.068	10.594	0.000	10.419	10	1.000
Z7	0.473	0.073	6.501	0.000	6.209	8	1.000
L3 =~							
Z8	1.000						
Z9	0.895	0.075	12.009	0.000	5.281	10	1.000
Z10	0.885	0.071	12.474	0.000	9.269	10	1.000
L4 =~							
Z11	1.000						
Z7	0.388	0.079	4.906	0.000	6.209	8	1.000
Z12	1.098	0.059	18.746	0.000	7.632	10	1.000
Z13	0.428	0.034	12.686	0.000	11.850	10	1.000
L1 ~							
Z1	3.828	0.315	12.142	0.000			
L2 ~							
L1	-0.716	0.092	-7.780	0.000	10.436	5	0.638
L3 ~							
L1	0.651	0.078	8.329	0.000	11.951	4	0.195
L4 ~							
L2	0.352	0.086	4.074	0.000	9.424	5	0.840
L3	-0.440	0.094	-4.688	0.000			

SOLUTION: Reisenzein (1986)

Cross-loading Model 2 Syntax:

```
model.cross.2 <- '
  L2 =~ Z5 + Z6 + Z7
  L1 =~ Z2 + Z3 + Z4
  L3 =~ Z8 + Z9 + Z10
  L4 =~ Z11 + Z12 + Z13 + Z7

  L4 ~ L2 + L3 + Z1
  L2 ~ L1
  L3 ~ L1
  L1 ~ Z1
 '
```

SOLUTION: Reisenzein (1986)

Fit Cross–Loading Model 2:

```
miive(model.cross.2, reisenzein1986, , sarg.adjust = "holm")
```

SOLUTION: Reisenzein (1986)

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
  L1 =~
    Z2                1.000
    Z3                1.047    0.080   13.041    0.000   12.085   10    1.000
    Z4                1.152    0.087   13.218    0.000   11.550   10    1.000
  L2 =~
    Z5                1.000
    Z6                0.724    0.068   10.594    0.000   10.419   10    1.000
    Z7                0.473    0.073    6.501    0.000    6.209    8    1.000
  L3 =~
    Z8                1.000
    Z9                0.895    0.075   12.009    0.000    5.281   10    1.000
    Z10               0.885    0.071   12.474    0.000    9.269   10    1.000
  L4 =~
    Z11               1.000
    Z7                0.388    0.079    4.906    0.000    6.209    8    1.000
    Z12               1.098    0.059   18.746    0.000    7.632   10    1.000
    Z13               0.428    0.034   12.686    0.000   11.850   10    1.000

  L1 ~
    Z1                3.828    0.315   12.142    0.000
  L2 ~
    L1               -0.716    0.092   -7.780    0.000   10.436    5    0.638
  L3 ~
    L1                0.651    0.078    8.329    0.000   11.951    4    0.195
  L4 ~
    Z1               -1.059    0.465   -2.274    0.023    5.330    4    1.000
    L2                0.255    0.092    2.769    0.006
    L3               -0.303    0.107   -2.827    0.005
```

SOLUTION: Reisenzein (1986)

Model	χ^2 (N = 138)	*df*	*p*
Null model	1,463.94	78	<.001
Model 1	86.58	61	<.020
Model 2 (with β_{HC} added)	85.95	60	<.016
Model 3 (with a correlation between the error terms of sympathy and anger added)	85.44	60	<.018
Model 4 (with both β_{HC} and the correlation between the errors of sympathy and anger added)	84.97	59	<.016
Model 5 (with β_{HE} added)	82.25	60	<.029
Model 5′ (= 5 with 2 correlated error terms)	67.28	58	≅.189
Model comparisons			
1 versus 2	0.63	1	.50 > *p* > .25
1 versus 3	1.14	1	.50 > *p* > .25
1 versus 4	1.61	2	.50 > *p* > .25
1 versus 5	4.33	1	<.05
1 versus null	1,377.36	17	<.001
5 versus null	1.69	18	<.001
5 versus 5′	14.97	2	<.001

Cross-loading Model 1: $\chi^2(60, N = 138) = 67.244$

Cross-loading Model 2: $\chi^2(59, N = 138) = 60.143$

EXTENSIONS

Bollen, Fisher, Giordano, Lilly, Luo, & Ye (forthcoming, *Psychological Methods)* has overview & extensive bibliography. Partial list below:

- **Categorical endogenous variables**
  - Bollen & Maydeu-Oliveres (2007)
  - Nestler (2012)
  - Jin, Luo, & Yang-Wallentin (2016)
  - Fisher & Bollen (2020)
  - Jin, Yang-Wallentin, Bollen (2021)
- **Multilevel modeling**
  - Giordano and Bollen (2020)

EXTENSIONS

- **Robustness**
  - Bollen, Gates & Fisher (2018)
  - Bollen (2020)
- **Interactions of latent variables**
  - Bollen (1995)
  - Bollen & Paxton (1998)
- **2nd Order growth curve models**
  - Nestler (2014)

EXTENSIONS

- **Higher order factor analysis**
  - Bollen & Biesanz (2002)
- **Specification error tests for nonlinearity and interactions**
  - Nestler (2015)
- **Testing dimensionality of measures**
  - Bollen (2011)
- **General Method of Moments estimator**
  - Bollen, Kolenikov, & Bauldry (2014)

Resources

lavaan Resources:

- http://lavaan.ugent.be/

MIIVsem Github

- https://github.com/zackfisher/MIIVsem
- readme file with examples

MIIVsem Planned Developments

- Missing data (normal or nonnormal distributions)
- General Method of Moments estimator
  - Permits subsets of equations to estimate
- Additional diagnostic tests of MIIVs
- Weak MIIV diagnostics

REFERENCES

Bollen, K. A. (1995). "Structural Equation Models That Are Nonlinear in Latent Variables: A Least-Squares Estimator." *Sociological Methodology* 25: 223–51.

Bollen, K.A., Fisher, Z., Giordano, M.L., Lilly, A., Luo, Lan, Ye, A. (forthcoming). An Introduction to Model Implied Instrumental Variables using Two Stage Least Squares (MIIV-2SLS) in Structural Equation Models (SEMs). *Psychological Methods*.

Bollen, K. A., and Paxton, P. (1998). "Interactions of Latent Variables in Structural Equation Models." *Structural Equation Modeling* 5 (3): 267–93.

Bollen, K. A. (2001). "Two-Stage Least Squares and Latent Variable Models: Simultaneous Estimation and Robustness to Misspecifications." In Structural Equation Modeling: Present and Future : A Festschrift in Honor of Karl Jöreskog, edited by Robert Cudeck, K. G. Jöreskog, and Dag Sörbom. Scientific Software International.

Bollen, K. A. and Biesanz, J. C. (2002). "A Note on a Two-Stage Least Squares Estimator for Higher-Order Factor Analyses."*Sociological Methods & Research* 30 (4): 568–79.

Bollen, K. A., & Bauer, D. J. (2004). Automating the Selection of Model-Implied Instrumental Variables. *Sociological Methods & Research*, *32*(4), 425–452.

Bollen, K. A. and Maydeu-Olivares, A. (2007a). "A Polychoric Instrumental Variable (PIV) Estimator for Structural Equation Models with Categorical Variables." *Psychometrika* 72 (3): 309–26.

REFERENCES

Bollen, K. A., Kirby, J. B., Curran, P. J., Paxton, P. M., & Chen, F. (2007b). Latent Variable Models Under Misspecification: Two-Stage Least Squares (2SLS) and Maximum Likelihood (ML) Estimators. *Sociological Methods & Research*, *36*(1), 48–86.

Bollen, K. A.(2011). "Evaluating Effect, Composite, and Causal Indicators in Structural Equation Models. "*MIS Quarterly* 35 (2): 359–72.

Bollen, K. A., Kolenikov, S. and Bauldry, S. (2014). "Model-Implied Instrumental Variable Generalized Method of Moments (MIIV-GMM) Estimators for Latent Variable Models. " *Psychometrika* 79 (1): 20–50.

Fisher, Z. and Bollen, K.A. (2020). An instrumental variable estimator for mixed indicators: Analytic derivatives and alternative parameterizations. *Psychometrika* 85(3), 660–683. https://doi.org/10.1007/s11336-020-09721-6

Fisher, Z. F., Bollen, K. A. Gates, K. and Rönkkö, M. (2017). MIIVsem: Model Implied Instrumental Variable (MIIV) Estimation of Structural Equation Models. R package Version 0.5.2.

Jin, S., Luo, H., and Yang-Wallentin, F. (2016). "A Simulation Study of Polychoric Instrumental Variable Estimation in Structural Equation Models. "*Structural Equation Modeling: A Multidisciplinary Journal* 23 (5): 680–94.

REFERENCES

Jin, S., Yang-Wallentin, F., and Bollen, K.A. (forthcoming). A unified model-implied instrumental variable approach for structural equation modeling with mixed variables. *Psychometrika*.

Kirby, J. B., and Bollen, K. A. (2009). "Using Instrumental Variable (IV) Tests to Evaluate Model Specification in Latent Variable Structural Equation Models." *Sociological Methodology* 39 (1): 327–55.

Nestler, S. (2013). "A Monte Carlo Study Comparing PIV, ULS and DWLS in the Estimation of Dichotomous Confirmatory Factor Analysis." *British Journal of Mathematical and Statistical Psychology* 66 (1): 127–43.

Nestler, S. (2014). "How the 2SLS/IV Estimator Can Handle Equality Constraints in Structural Equation Models: A System-of-Equations Approach." *British Journal of Mathematical and Statistical Psychology* 67 (2): 353–69.

Nestler, S. (2015). "Using Instrumental Variables to Estimate the Parameters in Unconditional and Conditional Second-Order Latent Growth Models." *Structural Equation Modeling: A Multidisciplinary Journal* 22 (3): 1–13.

Rosseel, Yves. (2012). "lavaan: An R Package for Structural Equation Modeling." *Journal of Statistical Software* 48 (2): 1–36.