MODEL IMPLIED INSTRUMENTAL VARIABLES (MIIVs): A NEW ORIENTATION TO STRUCTURAL EQUATION MODELING

Kenneth A Bollen

University of North Carolina at Chapel Hill

INTRODUCTION

Confusing Ideals with Reality

- Pure, ideal capitalism
  - Perfect competition
  - Optimal allocation of goods at optimal prices
  - Maximize utility ("happiness")
- Capitalism in practice
  - Markets dominated by few firms
  - Distortions in allocation of goods
  - Prices reflect lack of competition
  - Inequalities in utility

INTRODUCTION

Confusing Ideals with Reality

What does this have to do with modeling???

INTRODUCTION

Confusing Ideals with Reality

- System Wide Maximum Likelihood (ML)
  - Pure, ideal ML estimator properties
    - Consistent
    - Asymptotic unbiased
    - Asymptotic efficient
    - Asymptotic normality
    - Asymptotic standard errors
  - Fine Print
    - Correctly specified model
    - Multivariate normality
    - Sufficiently large sample

INTRODUCTION

Confusing Ideals with Reality

- System Wide Maximum Likelihood (ML)
  - Fine Print
    - Correctly specified model
    - Multivariate normality
    - Sufficiently large sample

INTRODUCTION

Confusing Ideals with Reality

- System Wide Maximum Likelihood (ML)
- SEM in reality with ML estimator
  - Approximate models
  - Biased and inconsistent estimator
  - No guarantee of asymptotic efficiency
  - No guarantee of accurate standard errors

INTRODUCTION

Approximate nature of SEMs

- Approximate = Misspecified
- Two forms of approximation
  - Distributional misspecification
    - nonnormal distributions
  - Structural misspecifications
    - Wrong model for relationships

INTRODUCTION

Structural misspecifications

- More serious problem than distributional misspecification
- Biased & inconsistent estimator of parameters
- Given approximate nature of models, ideal is to:
  - Detect where misspecification located
  - Prevent misspecification from spreading to parameter estimates in valid parts of model

INTRODUCTION

- Underidentified models with ML
  - 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

Maximum likelihood & system wide estimators

1. Negative impact of distributional misspecification
   - Significance tests inaccurate
2. Structural misspecifications effects can spread beyond bad parts of model

INTRODUCTION

Maximum likelihood & system wide estimators

3. Global tests of fit
   - Large N nearly always leads to significant chi square test given approximate nature of models
   - Locating source of problem difficult
     - Bad measurement model?
     - Bad latent variable model?
     - Modification index not always successful
4. Identified equations in underidentified models are not estimable

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

Purposes

1. Describe Model Implied Instrumental Variable (MIIV) estimator that meets these needs
2. Explain this approach to SEMs
3. Contrast it with system wide approach
4. Give current capabilities and future developments

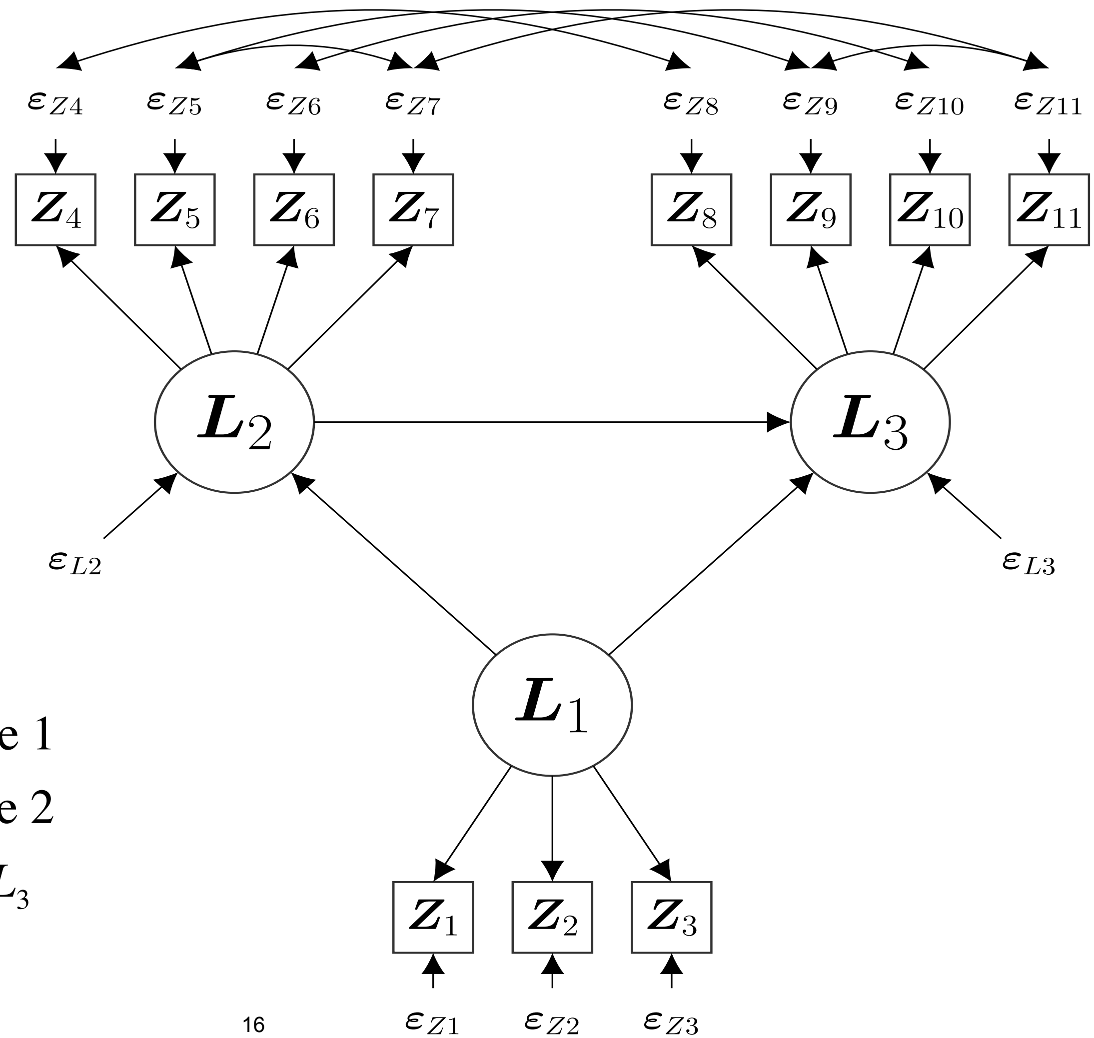
PRIMARY INGREDIENTS

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

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

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

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

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

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

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

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

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

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

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

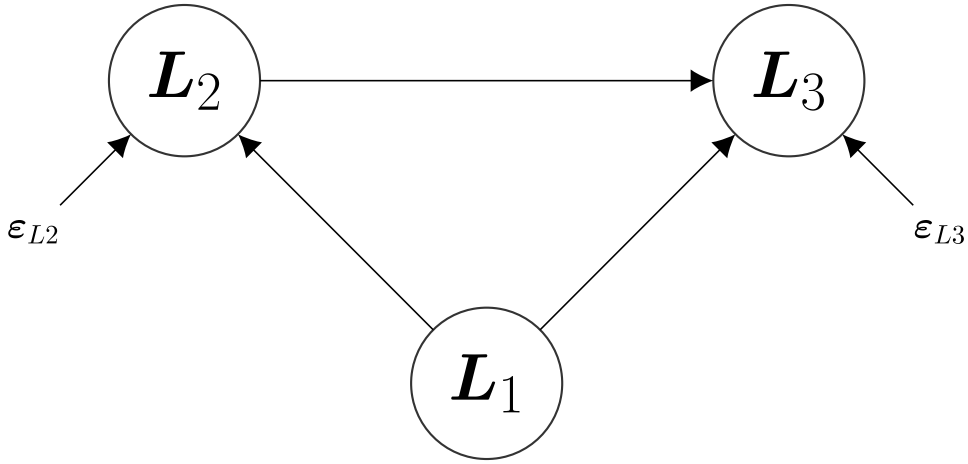
$$Z_{11} = \Lambda_{11,3}L_3 + \varepsilon_{z11}$$

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)



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.

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

3. Find Model Implied Instrumental Variables (MIIVs)
    - Key property of instruments is that they are uncorrelated with equation error
    - Typically, researchers search for instruments outside of variables already in model
    - 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)
   - General algorithm to find MIIVs (Bollen, 1996)
     - SAS: macro to implement in Bollen & Bauer (2004)
     - Stata: miivfind program in Bauldry (2014)
   - Expanded algorithm to non-standard models and lavaan (Rosseel, 2012) model syntax
     - R: MIIVsem (Fisher, Bollen, Gates & Rönkkö)
   - Though programs automatically find MIIVs, useful to illustrate process with example

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):

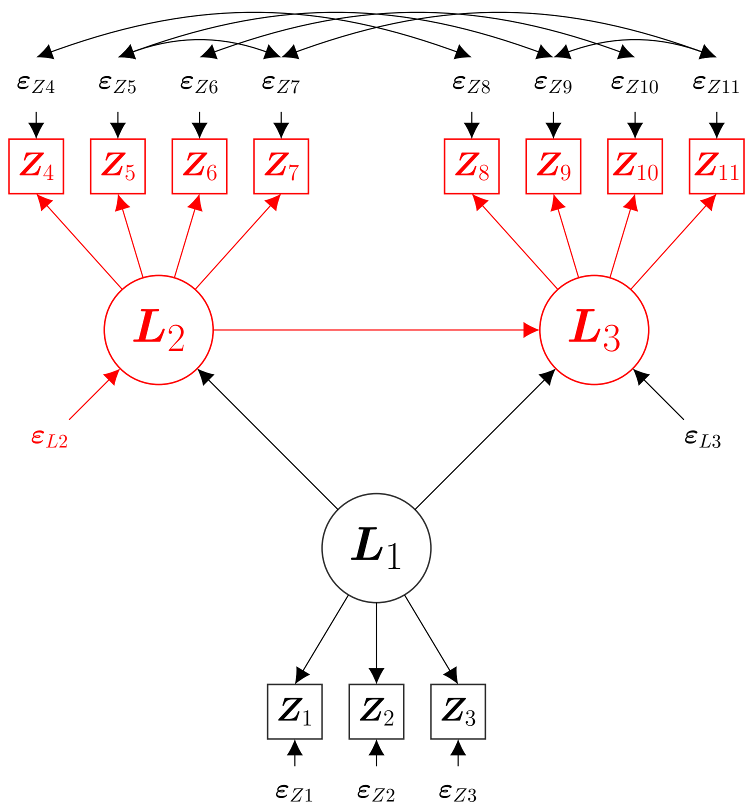
$$L_2 = \alpha_{L2} + B_{21}L_1 + \varepsilon_{L2} \quad \Rightarrow$$

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

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

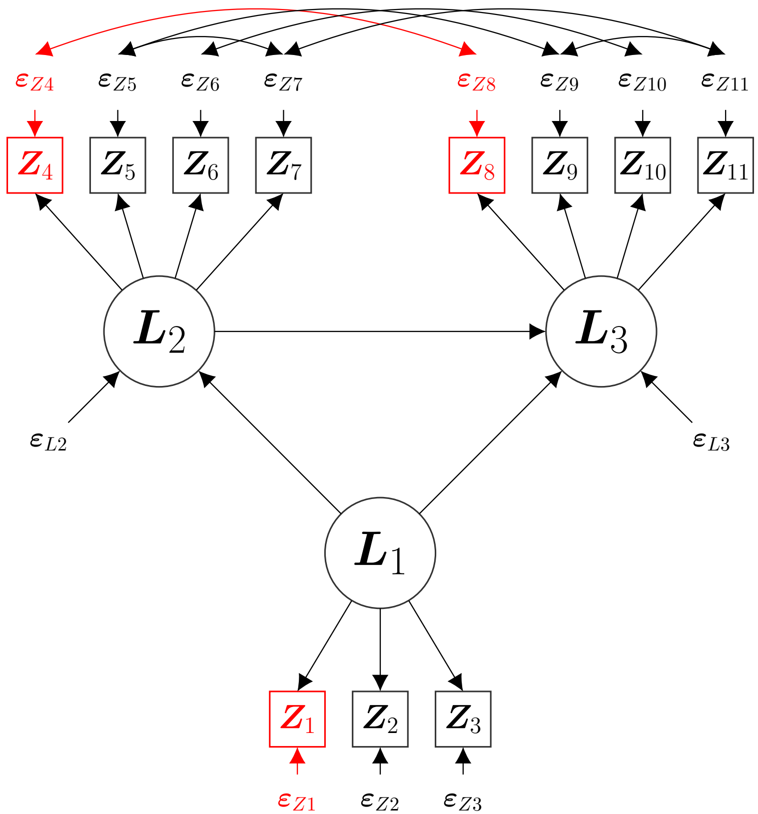
Find direct & indirect effects of ε_{L2}



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

PRIMARY INGREDIENTS

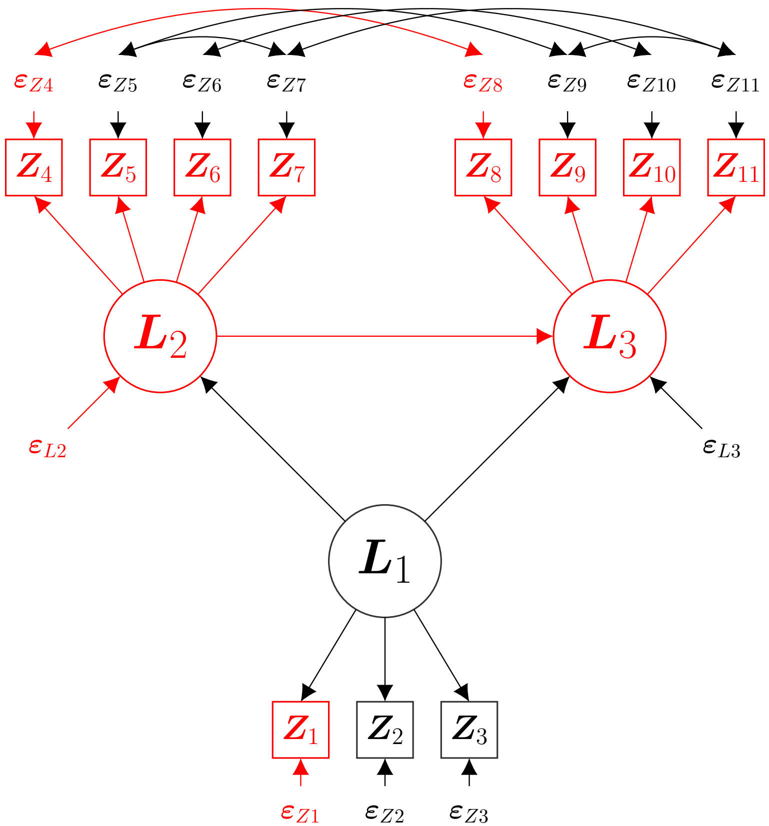
Find direct & indirect effects of $\varepsilon_{Z1}, \varepsilon_{Z4}$



Eliminates Z_1, Z_4, and Z_8 as MIIVs.

PRIMARY INGREDIENTS

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



Z_2, Z_3 only MIIVs.

PRIMARY INGREDIENTS

3. Find Model Implied Instrumental Variables (MIIVs)

The second latent variable equation, time 2 latent political democracy (L_3) regressed on time 1 latent political democracy (L_2) & industrialization (L_1):

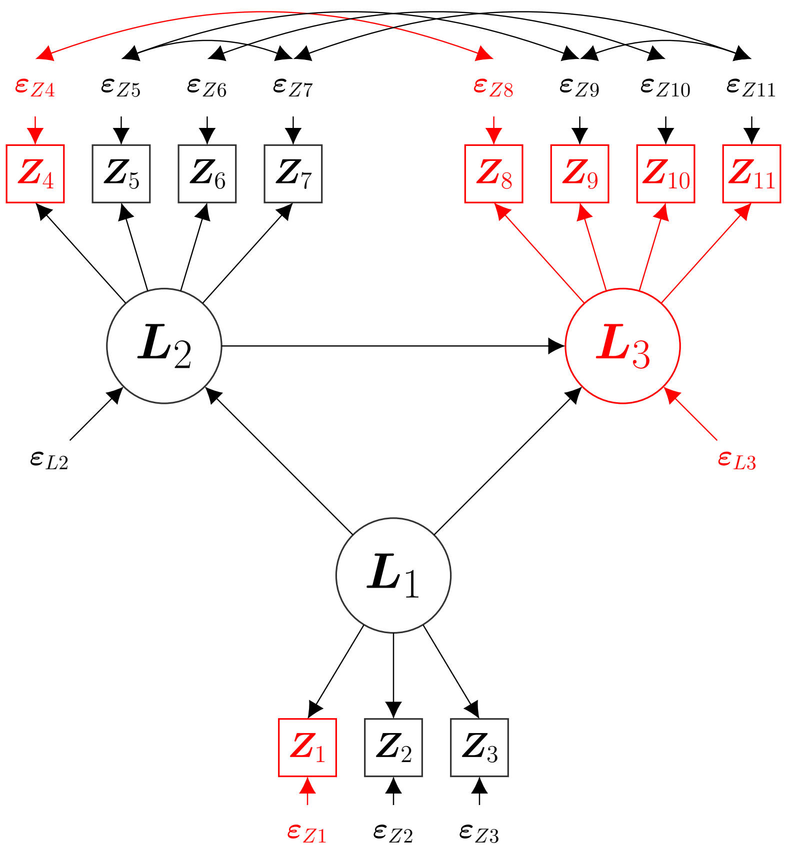
$$L_3 = \alpha_{L3} + B_{31}L_1 + B_{32}L_2 + \varepsilon_{L3} \quad \Rightarrow$$

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

Find direct & indirect effects of $\varepsilon_{Z1}, \varepsilon_{Z4}, \varepsilon_{Z8}, \varepsilon_{L3}$ on observed variables.

PRIMARY INGREDIENTS

Find direct & indirect effects of $\varepsilon_{Z1}, \varepsilon_{Z4}, \varepsilon_{Z8}, \varepsilon_{L3}$ on observed variables.



Z_2, Z_3, Z_5, Z_6, and Z_7 are MIIVs.

PRIMARY INGREDIENTS

3. Find Model Implied Instrumental Variables (MIIVs)
   - Previous slides illustrate finding MIIVs manually
   - General algorithm to find MIIVs (Bollen, 1996)
     - SAS: macro to implement in Bollen & Bauer (2004)
     - Stata: miivfind program in Bauldry (2014)
   - Expanded algorithm to non-standard models and lavaan (Rosseel, 2012) model syntax
     - R: MIIVsem (Fisher, Bollen, Gates & Rönkkö)

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):

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

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

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

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

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)

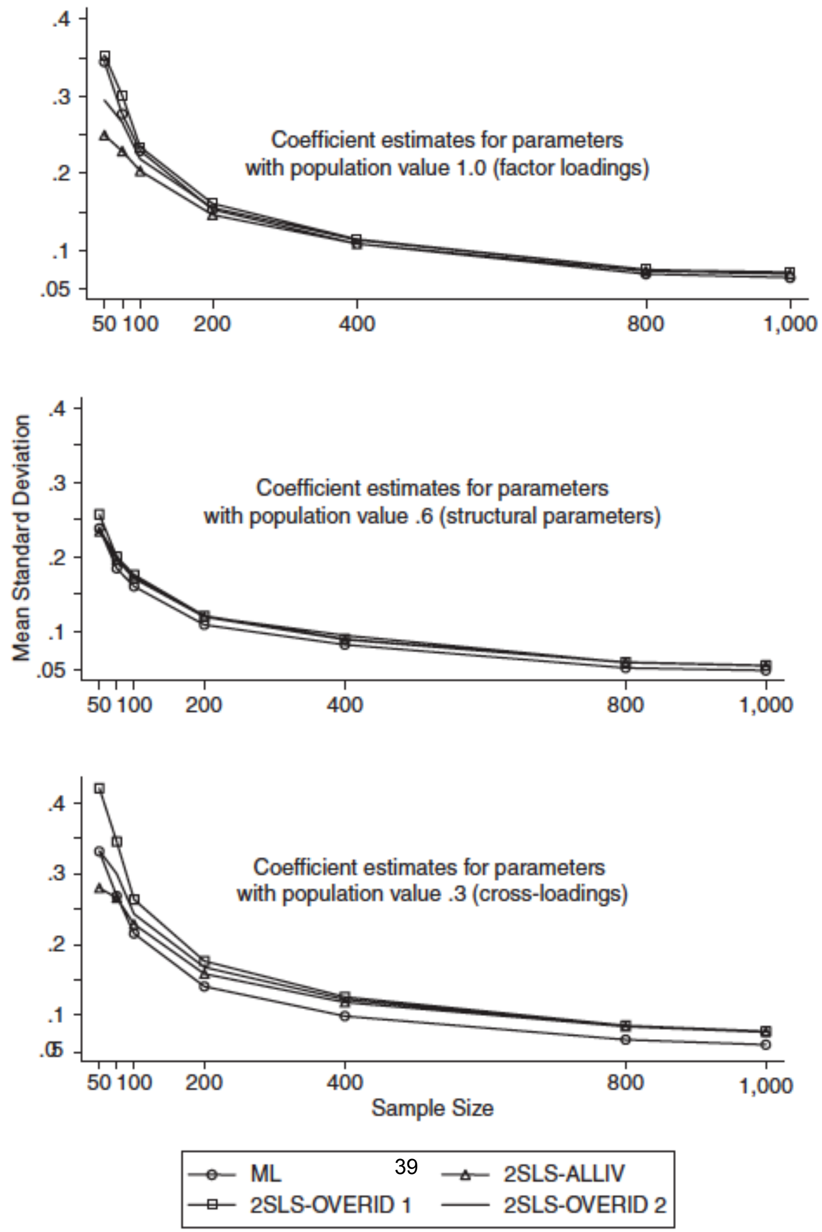
Illustration of ML and MIIV-2SLS simulation from Bollen, Kirby, Curran, Paxton, & Chen (2007b)

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

Normality

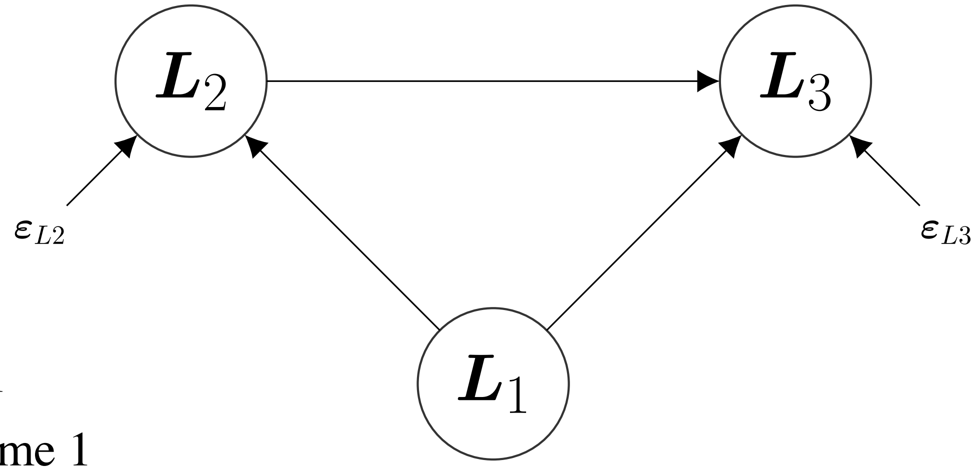
Correct specification

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

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

```
model <- '
     …
     L2 ~ L1
     L3 ~ L1 + L2
     …
'
```

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
 …
 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

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
```

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

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

```
model <- '
    …
    L2 ~ L1
    L3 ~ L1 + L2
    …
'
```

```
STRUCTURAL COEFFICIENTS:
                  Estimate  Std.Err  z-value  P(>|z|)   Sargan   df    P(Chi)
 …
 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

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
```

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

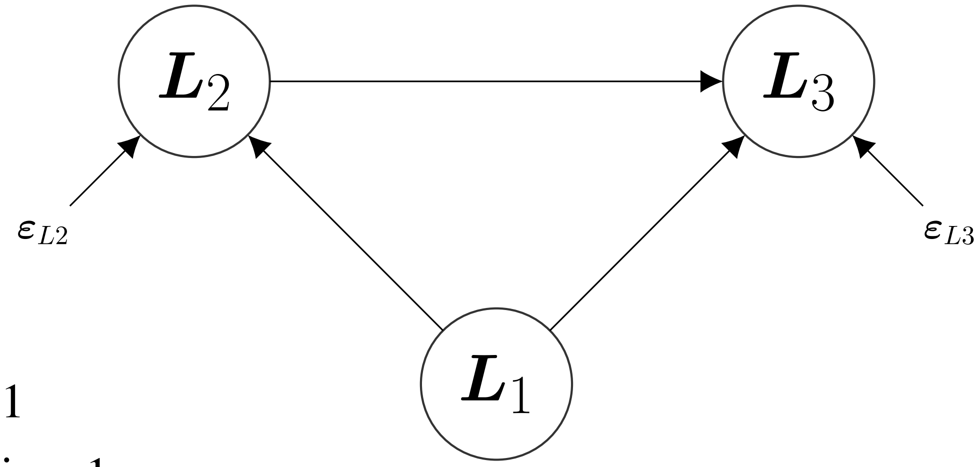
H_a: At least 1 MIIV correlates with error

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

PRIMARY INGREDIENTS

5. Tests each overidentified equation

- 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

5. Tests each overidentified equation

```
model <- ’
    …
    L2 ~ L1
    L3 ~ L1 + L2
    …
'
```

```
STRUCTURAL COEFFICIENTS:
                   Estimate  Std.Err  z-value  P(>|z|)   Sargan   df   P(Chi)
 …
 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

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
```

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

ROBUSTNESS

2. Structural misspecification robustness

- omitted paths
- omitted variables
- wrong number of dimensions

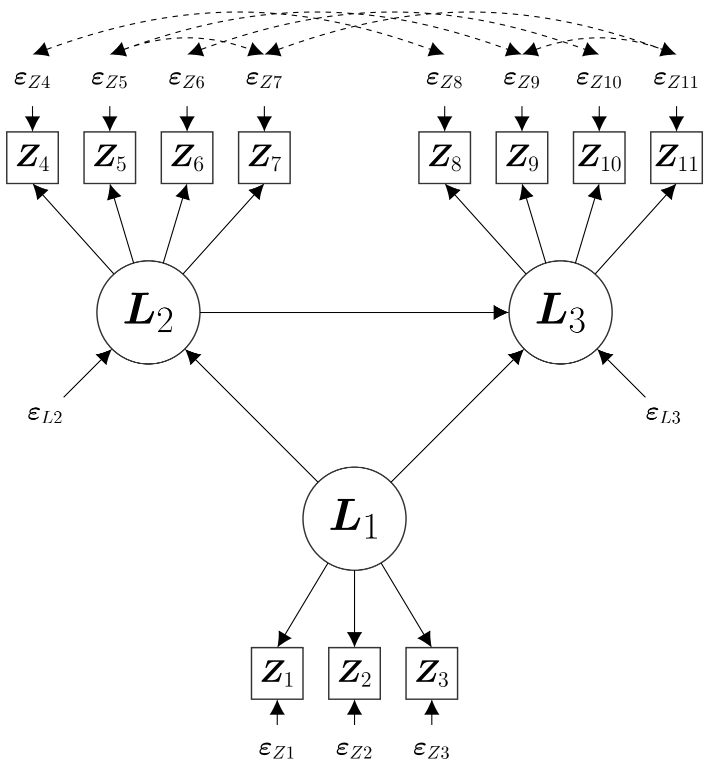
Bollen (2001): Suppose that for the j^{th} equation in the correctly specified model, the model implied IVs are in a matrix $\mathbf{V}_j$. The 2SLS estimator of the coefficients is robust for any misspecification in other equations under two conditions:

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

Suppose "true" model has dashed & solid lines, what happens when only solid line model assumed?



STRUCTURAL COEFFICIENTS:

	True Model	Omitting Correlated Errors
L2 ~		
L1	1.261	1.261
L3 ~		
L1	1.123	1.123
L2	0.724	0.724

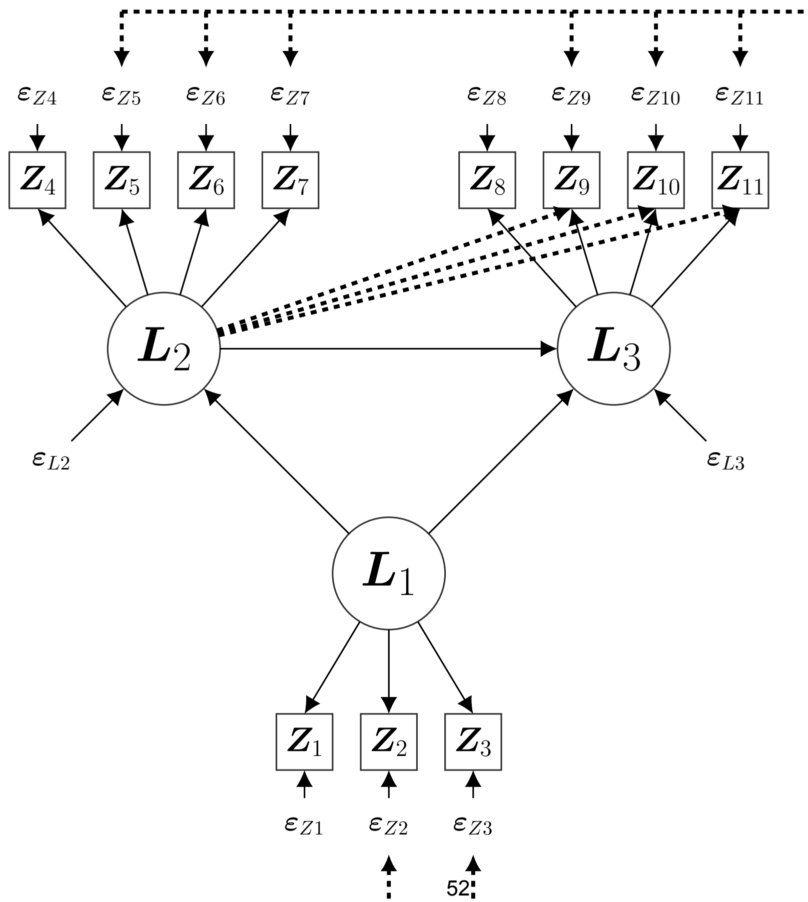
INTERCEPTS:

L2	-0.909	-0.909
L3	-4.498	-4.498

ROBUSTNESS

2. Structural misspecification robustness

   "True" model has dashed & solid lines



ROBUSTNESS

2. Structural misspecification robustness

STRUCTURAL COEFFICIENTS:

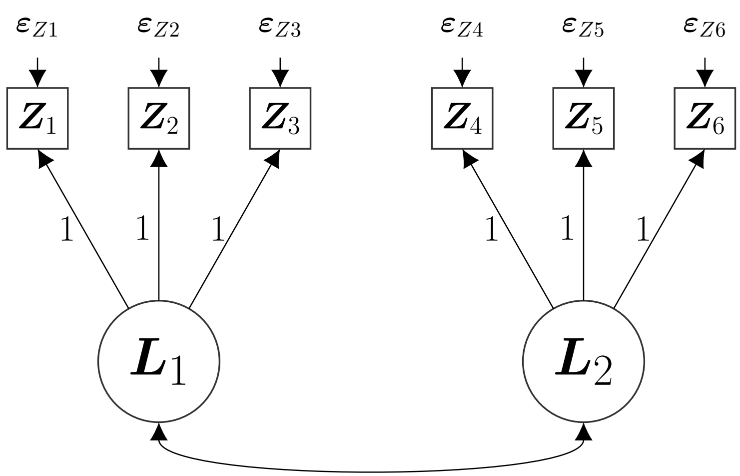
	True Model	Omit 6 Corr. Errors	Omit 28 Corr. Err. and 3 Λs
L2 ~			
L1	1.261	1.261	1.261
L3 ~			
L1	1.123	1.123	1.123
L2	0.724	0.724	0.724

INTERCEPTS:

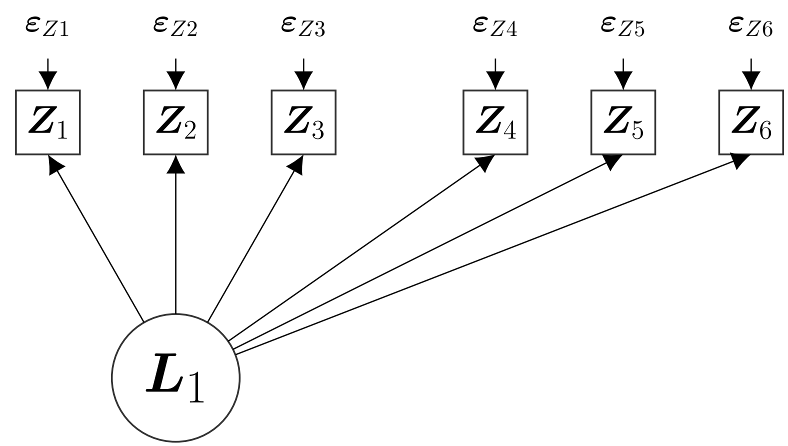
L2	-0.909	-0.909	-0.909
L3	-4.498	-4.498	-4.498

ROBUSTNESS

Correct Model

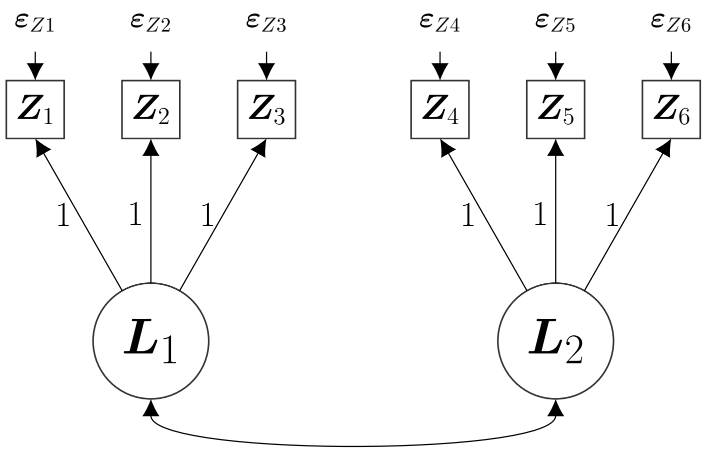


Incorrect Model

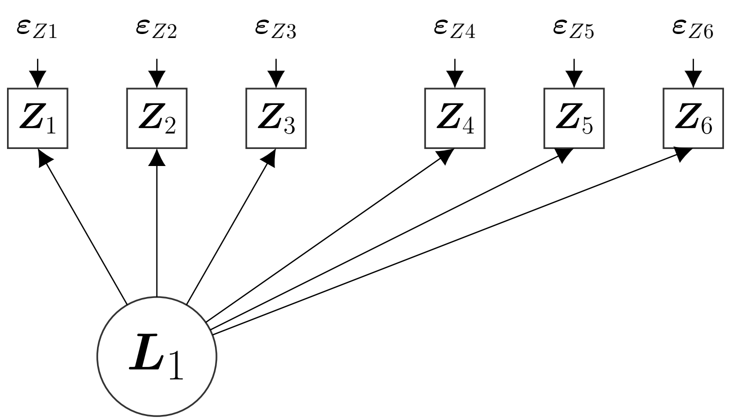


ROBUSTNESS

Correct Model



Incorrect Model



STRUCTURAL COEFFICIENTS:

	Two Factor Model	One Factor Model
L1 =~		
Z1	1.000	1.000
Z2	**1.034**	**1.034**
Z3	**0.964**	**0.964**
Z4	1.000	0.363
Z5	1.043	0.359
Z6	1.144	0.405

ROBUSTNESS

SARGAN TEST (p-value):

	Two Factor Model			One Factor Model		
	Sargan	df	P(Chi)	Sargan	df	P(Chi)
L1 =~						
Z1						
Z2	0.555	3	0.907	0.555	3	0.907
Z3	1.048	3	0.790	1.048	3	0.790
Z4				**246.738**	**3**	**0.000**
Z5	0.247	3	0.970	**263.962**	3	**0.000**
Z6	0.858	3	0.836	**273.846**	3	**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.

EXTENSIONS

- Categorical endogenous variables
  - Bollen & Maydeu-Oliveres (2007a)
  - Nestler (2012)
  - Jin, Luo, & Yang-Wallentin (2016)
- 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 tests for nonlinearity and interactions
  - Nestler (2015)
- Model specification tests
  - Kirby & Bollen (2009)
- Testing dimensionality of measures
  - Bollen (2011)
- General Method of Moments estimator
  - Bollen, Kolenikov, & Bauldry (2014)

EXTENSIONS

- Software
  - Finding MIIVs
    - Bollen & Bauer (2004) in SAS
    - Bauldry (2014) miivfind in Stata
    - Fisher, Bollen, Gates & Rönkkö MIIVsem in R

EXTENSIONS

- Software
  - MIIVsem [Fisher, Bollen, Gates & Rönkkö]
    - Designed for MIIV approach
    - Finds MIIVs
    - Covariance based input allowed
    - MIIV-2SLS estimator implemented
    - Sargan test statistic for overidentified equations
    - Equality restrictions and Wald tests
    - Bootstrap options
    - Uses lavaan (Rosseel, 2012) model syntax
    - Categorical endogenous variables modeled

EXTENSIONS

- Software
  - MIIVsem [Fisher, Bollen, Gates & Rönkkö]
    - Features under development
      - Missing data
      - General Method of Moment estimator (MIIV-GMM)
      - Lagrangian multiplier tests
      - Weak instrument diagnostics

CONCLUSIONS

- SEM is dominated by estimators that assume perfection while we simultaneously preach that models are approximations
  - Optimal properties of ML called into question
    - Claims of consistency, efficiency, etc. no longer supported

CONCLUSIONS

- Approximation = structural misspecifications
  - Desirable to distinguish good from bad parts of model
    - Suggest need for local tests
  - Want estimator less likely to spread bias
    - Suggest need for estimator with more robustness to structural misspecifications
  - Bonus if estimator "distribution free"

CONCLUSIONS

- MIIV-2SLS better satisfies the realities of approximate models
  - Each overidentified equation has an overidentification test
  - Less likely to spread bias from structural misspecifications through system
  - Asymptotic distribution free estimator

CONCLUSIONS

- Future research needs for MIIV-2SLS
  - Clarify robustness conditions
  - Optimal selection of MIIVs when there are many
  - Empirical methods to respecify poorly fit models
  - Further understand when MIIV-2SLS performs best and worse
    - E.g., Bollen et al. (2007b) found that at small Ns, best not to use large # of MIIVs, but matters less for large Ns

CONCLUSIONS

- SUMMARY OF MIIV APPROACH
  - We need to match our methods to the approximate nature of our models

"Specify Globally, Estimate and Test Locally"

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., 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. 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.

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.

REFERENCES

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.