

Structural Equation Modelling & Causal Inference

Set-1: General Introduction to Structural Equation Modelling (SEM)

Ozan Aksoy

UCL Social Research Institute & NCRM
University College London

November 6, 2023

Topics

- 1 Logistics of this module
- 2 Overview of SEM
- 3 Univariate regression
- 4 Multivariate regression
- 5 Recursive path models
- 6 Model fit
- 7 Imposing and testing constraints
- 8 Complicating models
- 9 Non-recursive path models
- 10 Identification

Course overview

3/42

- Set 1: basics
 - Basics of Structural Equation Modelling (SEM)
 - Structural (regression type) models
 - Measurement models (Confirmatory Factor Analysis)
 - Multiple group analysis and measurement invariance
- Set 2: longitudinal
 - Measurement invariance across time
 - Multi-trait multi-method models
 - Latent-curve models
 - Full Information Maximum Likelihood Estimation
- Set 3: causal analysis with SEM
 - SEM versus DAGs
 - Fixed versus random effects models
 - Cross-lagged panel models with fixed effects
 - Instrumental variable models

Plan

4/42

- Online lecture (10am-1pm)–Introduces the topics
- Computer practicals (2pm-5pm)–Hands-on exercises with supervision
- All material are at: <https://github.com/aksoyundan/SEM>
- Main software: R and RStudio
- Main R packages: lavaan and semTools and some others
- Install R, RStudio, lavaan, semTools etc. at lunch break if you haven't already:
<https://rstudio-education.github.io/hopr/starting.html>

Structural equation models

5/42

- Integration of regression models, path models, simultaneous equations, and factor analysis
- SEM may include **observed** (manifest) and **unobserved** (latent) variables
 - E.g., factors are unobserved/latent variables measured with observed indicators
 - E.g., social class is latent, income is observed
 - E.g., religiosity is latent, frequency of prayer, church visit etc. are observed
- Loosely, an (observed or latent) variable is called
 - **endogenous** if DV (maybe IV in other regressions): it is determined within “the system”
 - **exogenous** if only IV: it is determined outside “the system”
- **Errors** (residuals, disturbances) are exogenous latent variables
- SEM may include **covariance** and **mean** structures

Structural equation models (cont)

6/42

Equivalent names for rich class of linear models

- structural equation model (SEM)
- covariance structure analysis (CSA)
- linear structural relations (LISREL) ...

Modern extensions allow

- non-normal distributed continuous data
- generalized linear relations to deal with categorical (ordinal, binary) and nominal DVs
- limited forms of nonlinear regression (interactions, quadratic effects, ...)
- multilevel structures: the relations between variables at level 1 (e.g., individual) vary at level 2 (e.g., context)
- discrete latent variables (model-based clustering)
- Bayesian SEM
- Generalized latent variable modelling

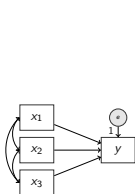
SEM diagrams

7/42

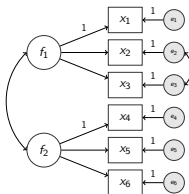
- Models may be denoted by a diagram
 - **rectangular box** for observed (manifest) variable
 - **ellipse** for unobserved (latent) variable, including errors
 - causal link: (one-sided) **arrow** from x to y variable
 - noncausal relation: **double pointed arrow** between two variables (usually: between two errors or two exogenous variables)
 - sometimes: intercepts represented by **triangles** (I don't do it)
- Some software (Amos, Stata, ...) allow model specification by drawing a *diagram*
- Nice for novices and simple models
- But ... infeasible for more complicated model
- Convenient communication of models/results
 - use **Tikz** in \LaTeX

Some examples of SEM diagrams (Tikz)

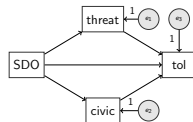
8/42



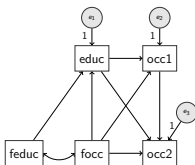
regression model



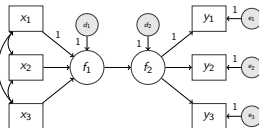
2 factor CFA with correlated factors



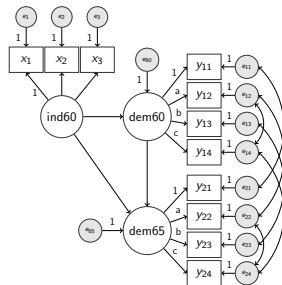
path model



Blau-Duncan path model



MIMIC model



Bollen dynamic model

Resources to learn about structural equation models

9/42

Books

- Kline 2023 5th. Principles and Practice of Structural Equation Modeling.
- Hancock Mueller 2013. Structural Equation Modeling. A Second Course.
- Bollen Curran 2006. Latent Curve Models. A SEM Perspective
- Hox et al 2017 3ed. Multilevel Analysis. Techniques and Applications.

Websites

- <http://davidakenny.net/kenny.htm>
- <http://www2.gsu.edu/~mkteer/index.html>
- <https://www.guilford.com/kline-materials>

Linear regression

10/42

Regression of continuous y on continuous predictors x_1 and x_2

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + e_i$$

$e_i \sim \text{Normal}(0, \sigma^2)$ (independent homoskedastic residuals)

$\text{cov}(x_j, e) = 0$ ("exogeneity of x_j ")

In R- lavaan,

- a regression is written as $y \sim \text{xlist};$
- intercept β_0 , coefficients β_j , and error term e_i are *implicit*

_____ simple regression _____

```

1 x1 <- rnorm(100, mean = 0, sd = 2)
2 x2 <- 0.3*x1 + rnorm(100, mean = 0, sd = 2)
3 y  <- 0.2 + 0.7*x1 + 1.2*x2 + rnorm(100, mean = 0, sd = 1)
4 d  <- as.data.frame(cbind(y, x1, x2))
5 library(lavaan)
6 mymodel <- 'y ~ x1 + x2
7           y ~ 1' # otherwise intercept is suppressed
8 fit <- sem(mymodel, data=d)
9 summary(fit, standardized = TRUE)
```

Regression analysis: Output

11/42

```

1 lavaan (0.5-23.1097) converged normally after 19 iterations
2   Number of observations              100
3 [omitted for brevity]
4 Regressions:
5
6           Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
7   y ~
8     x1             0.695    0.051   13.699    0.000    0.695    0.515
9     x2             1.069    0.062   17.197    0.000    1.069    0.646
10
11 Intercepts:
12
13           Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
14   .y              0.172    0.104    1.656    0.098    0.172    0.060
15
16 Variances:
17
18           Estimate  Std.Err  z-value  P(>|z|)  Std.lv  Std.all
19   .y              1.072    0.152    7.071    0.000    1.072    0.130

```

- Thus: $E(Y) = 0.2 + 0.7 \cdot X_1 + 1.1 \cdot X_2$, $\text{var}(e) = 1.1$
- Under H_0 : $b_x = 0$: $\hat{b}_x / \hat{\text{se}}(\hat{b}_x)$ approximately standard normal
- `standardized = TRUE` produces standardized solution; and $R^2 =$ 1-standardized residual variance.

R-lavaan operators

12/42

formula type	operator	mnemonic
latent variable definition	$=\sim$	is measured by
regression	\sim	is regressed on
(residual) (co)variance	$\sim\sim$	is correlated with
intercept	~ 1	intercept
new parameter	$:=$	defined as

Multivariate models

13/42

Often need to analyze multiple responses simultaneously:

- multiple dimensions of problem behavior (aggression, depression, ...)
- interpersonal trust, measured at different time points
- sex role attitudes of husbands and wives
- educational attainment, occupational status, income ...

Distinguish variables

- variable is **only y**
- variable is **only x**
- variable is **both y and x** (intermediate, mediator, ...)

Multivariate and seemingly unrelated regression

14/42

Models and assumptions applicable if **no** variable is **both y and x**

- MVREG/MANOVA: same x variables predicting the y's
- SUREG: separate spec (maybe overlap) of x's predicting the y's
- an y is not x for another y (otherwise: path, simeqns)
- errors across y's may covary
- better approach than fitting separate models per dv

Formal model specification of SUREG (and $\text{cov}(x,e)=0$):

$$\begin{aligned} y_{i1} &= \beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2} & + e_{i1} & \quad \text{var}(e_{ij}) = \sigma_j^2 \\ y_{i2} &= \beta_{20} + \beta_{21}x_{i1} & + \beta_{23}x_{i3} + e_{i2} & \quad \text{cov}(e_{i1}, e_{i2}) = \sigma_{12} \end{aligned}$$

In R:

- may combine different types of regression (e.g. binary-continuous)
- `y1 ~ y2`; specifies that covariance of $e.y1$ and $e.y2$. is a free parameter

MVREG-example

15/42

mvreg1

```

1 library(lavaan)
2 model2 <- 'y1 ~ x1 + x2
3           y2 ~ x1 + x2
4           y1 ~~ y2' # adds cov(e.y1,e.y2), default in R
5 fit2 <- sem(model2, data=d)
6 summary(fit2)

```

mvreg2

```

1 model3 <- 'y1 ~ x1 + x2
2           y2 ~ x1 + x2
3           y1 ~~ 0*y2' # removes cov(e.y1,e.y2) from model
4 fit3 <- sem(model3, data=d)
5 summary(fit3)

```

Recursive path model; direct and indirect effects

16/42

Formal model specification with mediator:

$$\begin{aligned} y_{i1} &= \beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2} + \gamma_{12}y_{i2} + e_{i1} & \text{var}(e_{ij}) &= \sigma_j^2 \\ y_{i2} &= \beta_{20} + \beta_{21}x_{i1} + \beta_{23}x_{i3} + e_{i2} & \text{cov}(e_{i1}, e_{i2}) &= 0 \end{aligned}$$

Reduced form by substitution

$$\begin{aligned} y_{i1} &= \beta_{10} + \beta_{11}x_{i1} + \beta_{12}x_{i2} + \gamma_{12}(\beta_{20} + \beta_{21}x_{i1} + \beta_{23}x_{i3} + e_{i2}) + e_{i1} \\ &= (\beta_{10} + \gamma_{12}\beta_{20}) + (\beta_{11} + \gamma_{12}\beta_{21})x_{i1} + \beta_{12}x_{i2} + \gamma_{12}\beta_{23}x_{i3} + (e_{i1} + \gamma_{12}e_{i2}) \end{aligned}$$

In words:

- x_1 affects y_1 directly, and indirectly via y_2
- total effect of x_1 = direct (β_{11}) + indirect effect ($\gamma_{12}\beta_{21}$)
- General case: indirect effect of x on y is **sum** over all possible paths from x to y of **product** of path coefficients
- Below: In (nonrecursive) model with causal cycles, *infinitely* many possible paths, but *sum* mathematically well-defined
- SEs of indirect/total effects by delta method (or bootstrapsap)

Path model in R

17/42

Blau & Duncan

```
1 bd_low <- '
2 1.0000
3 0.5160 1.0000
4 0.4530 0.4380 1.0000
5 0.3320 0.4170 0.5380 1.0000
6 0.3220 0.4050 0.5960 0.5410 1.0000'
7 bd.corr <- getCov(bd_low, names = c("faed", "faocc",
8                                     "educ", "occ1", "occ2"))
9 m.bd <- 'educ ~ a*faed + b*faocc
10         occ1 ~ c*educ + d*faocc
11         ac   := b*c
12         total := d + (b*c)'
13 fit.bd <- sem(m.bd, sample.cov = bd.corr, sample.nobs = 20700)
14 summary(fit.bd)
```

Path model in R—selected output

18/42

```

1 lavaan (0.5-23.1097) converged normally after 12 iterations
2 ...
3 Regressions:
4
5           Estimate Std.Err z-value P(>|z|)
6   educ ~
7     faed      (a)   0.309   0.007  44.383   0.000
8     faocc     (b)   0.278   0.007  39.937   0.000
9   occ1 ~
10    educ      (c)   0.440   0.006  69.488   0.000
11    faocc     (d)   0.224   0.006  35.463   0.000
12 Variances:
13
14           Estimate Std.Err z-value P(>|z|)
15   .educ      0.738   0.007 101.735   0.000
16   .occ1      0.670   0.007 101.735   0.000
17
18 Defined Parameters:
19
20           Estimate Std.Err z-value P(>|z|)
21   ac          0.122   0.004  34.625   0.000
22   total       0.347   0.007  53.027   0.000

```

Note: $0.122 = 0.278 (FAOCC \rightarrow EDUC) \times 0.440 (EDUC \rightarrow OCC1)$

Indirect effects vs. mediation: causality caveat!!

19/42

- Mediation refers to **causal** hypothesis
- Mediation involves indirect effect
- But not all indirect effects signal mediation
- Mediation should be used sparingly, more specifically...
- **Sequential ignorability assumption** has to be met for appropriate tests of mediation:
 - $X \rightarrow M \rightarrow Y; X \rightarrow Y$
 - X has to be exogenous to M and Y (no confounding between X, M, and Y)
 - M has to be exogenous to Y (no confounding between M and Y)
 - No interaction between X and M (can be relaxed)
- If interested: read on **causal mediation analysis**

Path model: multiple indirect effects

20/42

In the specification below, x1 influences y1 in three ways

- directly
- indirectly via y2
- indirectly via y3

total effect = direct effect + sum of indirect effects

```

1 M <- 'y1 ~ a*y2 + b*y3 + c*x1 + x2
2     y2 ~ x3 + d*x1
3     y3 ~ x4 + e*x1
4     id1 := d*a # via y2
5     id2 := e*b # via y3
6     tot := c + d*a + e*b # total effect
    
```

✚ Equivalent models

21/42

- Different models with **same fit**, but very **different interpretation**
- Often arrows may be reversed, replaced by mutual influence, or freeing covariance among residuals!
- Examples of 3 equivalent models:

$$(1) \quad x \rightarrow y \rightarrow z$$

$$(2) \quad x \leftarrow y \leftarrow z$$

$$(3) \quad x \leftarrow y \rightarrow z$$

- Often DOZENS-HUNDREDS of equivalent models; Theory is unavoidable!
- Rarely addressed in application; no good software tools to sensitize researchers

Assessing Fit: Model Chi-square

22/42

- **LRtest** of model vs **saturated** model
 - Hope: not significant, test of **exact fit hypothesis**
 - $df = \text{sample moments} - \text{number of free parameters}$
 - sensitive to multivariate normality
 - With some estimators (eg MLM, ...), tests adjusted for nonnormality
 - sensitive to sample size: complex (maybe stupid) model not rejected in small samples
 - reasonable but untrue (all!) models rejected in big samples
- **LRtest** of model vs **base** model
 - Hope: highly significant
 - common base model: independence of variables
 - compare null model is regression
 - base nearly always fits very badly (why?)
 - base is used to evaluate quality of model
hardly an accomplishment to improve on the silly, use a more worthy opponent: a more informative baseline

Example: fit measures of Blau-Duncan model

23/42

Blau & Duncan fit (selection)

```

1  lavaan (0.5-23.1097) converged normally after 12 iterations
2    Number of observations                    20700
3    Estimator                                ML
4    Minimum Function Test Statistic          13.361
5    Degrees of freedom                       1
6    P-value (Chi-square)                     0.000
7  Model test baseline model:
8    Minimum Function Test Statistic          14598.385
9    Degrees of freedom                       5
10   P-value                                  0.000
11 User model versus baseline model:
12   Comparative Fit Index (CFI)              0.999
13 Root Mean Square Error of Approximation:
14   RMSEA                                    0.024
15   90 Percent Confidence Interval            0.014 0.037
16   P-value RMSEA <= 0.05                    1.000
17 Standardized Root Mean Square Residual:
18   SRMR                                      0.005
  
```

Comparative fit indices

24/42

- compare fit of model with fit of baseline
- recall: baseline is unworthy opponent
- many indices in literature, R reports TLI and CFI
- Comparative Fit Index (good CFI > 0.90, better: > 0.95)
- Do not rely on strict cutoff values!

$$\text{CFI} = 1 - \frac{\max(\chi^2_{\text{model}} - \text{df}_{\text{model}}, 0)}{\max(\chi^2_{\text{base}} - \text{df}_{\text{base}}, 0)}$$

Absolute fit indices

25/42

- Root Mean Square Error Approximation

$$\text{RMSEA} = \sqrt{\frac{\max(\chi^2_{\text{model}} - \text{df}_{\text{model}}, 0)}{\text{df}_{\text{model}}(N - 1)}}$$

- recall: if model fits, $E(\chi^2_{\text{model}}) = \text{df}$
- $\text{RMSEA} < .08$ (reasonable fit); $\text{RMSEA} < .05$ (good fit)
- R reports 90CI and $\text{PCLOSE} = P(\text{RMSEA} < .05)$.
- Good fit: $\text{hi}(90\text{CI}) < .05$ and PCLOSE close to 1.
- Root Mean of squared standardized Residuals ($\text{SRMR} < 0.05$)

$$\text{SRMR} = \sqrt{\frac{1}{ns} \left(\sum_{ij} \left(\frac{S_{ij} - \hat{\Sigma}_{ij}}{S_{ij}} \right)^2 + \sum_i \left(\frac{m_i - \hat{\mu}_i}{m_i} \right)^2 \right)}$$

Here: ns is number of sample moments, S_{ij} and m_i are sample moments, $\hat{\Sigma}_{ij}$ and $\hat{\mu}_i$ are fitted moments

- If χ^2 test is rejected, check residual moments (obs-fitted covs)

Checking residual moments: Blau & Duncan example 26/42

```

1 > resid(fit.bd, type="standardized")
2 $type
3 [1] "standardized"
4
5 $cov
6      educ  occ1  faed  faocc
7 educ  0.000
8 occ1  0.000 0.000
9 faed  0.000 2.477 0.000
10 faocc 0.000 0.000 0.000 0.000
11
12 $mean$
13      educ  occ1  faed  faocc
14      0      0      0      0
  
```

✚ Information criteria

27/42

Used to compare *nonnested* models

- Akaike's Information Criterion

$$\text{AIC} = C + 2q$$

- Bayesian/Schwarz Information Criterion

$$\text{BIC} = C + \ln(N) * q$$

- C = LR-test-statistic comparing *model* with *saturated*
- q = # parameters
- Lower AIC or BIC is better

Minimum set of fit statistics

28/42

- Model χ^2 , df, p-value (exact-fit hypothesis)
- RMSEA with 90% CI (close-fit hypothesis)
- CFI
- SRMR

If exact-fit hypothesis is rejected, analyze the residual correlation matrix, using data (residuals, modification indices etc.) modify the model in a theory guided and transparent way

Model selection and model improvement

29/42

- strike a better balance between fit and complexity:
 - simplify model if possible (model trimming) – little loss in fit
 - complicate model if necessary (model building) – major improvement in fit
- simplifications to consider
 - constrain parameters to 0: remove arrows
 - equate parameters (eg controls have same effects on different yvars or across groups)
 - remove items in scale
- complications to consider (use modification indices)
 - free parameters
 - relax equality constraints

Fixing, freeing, initializing parameters

30/42

Fix parameters to specific values: **VALUE*x**

- to simplify the model (eg, fix covariance to 0)
- to identify the model
- to avoid inadmissible solution (eg, fix a variance to 0 to circumvent a negative estimated value)
- substantive reasons

Setting a parameter free: **NA*x** or adding the parameter explicitly

- free a covariance that would otherwise be assumed to be 0
- in CFA: free loading of item (next weeks)
- relax cross-group constraints (next weeks)

Initialize parameters to reasonable values: **start(0.8)*x2**

- Iterative fitting requires good starting values
- R- lavaan is often ok, sometimes needs help

Fixing, freeing, and initializing parameters—example

31/42

```

1
2 model <- 'y1 ~ 1*x1 + x3          #b(x1) fixed to 1
3           y2 ~    x1 + x3
4           y1 ~ 0*1                #intercept fixed to 0
5           y1 ~~ start(0.12)*y1    #initial value error variance
6           y1 ~~ y2'               #explicit frees cov(e.y1,e.y2)
7 fit <- sem(model, data=d)

```

Imposing equality constraints—2 methods

32/42

```

1 model1 <- 'y1 ~ p*x1 + p*x2 # === METHOD 1 ===
2                                     #b(x2) and b(x2) labeled p
3           y2 ~ x1
4           y1 ~~ a*y1             #var(e.y1) and var(e.y2)
5           y2 ~~ a*y2'           #labeled a hence equal

```

```

1 mod1    <- 'y1 ~ a*x1 + b*x2 # === METHOD 2 ===
2           y2 ~ x1
3           y1 ~~ c*y1
4           y2 ~~ d*y2
5           a == b                 #constrains a to be equal to b
6           c == d^2'             #non-linear constraints possible

```

Explicit constraints are more general, may be nonlinear:

$p_1 = p_2/2$, $p_3 = p_1 + p_2$, $p_1 * p_2 = 1$, $p_1 > p_3$, etc (but: R may have difficulty fitting the model ...)

Methods may be mixed

Testing constraints (Wald-style)

33/42

In regression of y on x_1 x_2 x_3 , two common tests:

- test $b(x_1)=0$ and $b(x_2)=0$
- test $b(x_1)=b(x_2)$

In R-lavaan:

- give labels (cannot start with a number) to parameters,
- or extract default labels used by lavaan (`coef(fit)`)
- Wald test: use `lavTestWald()` with (non)linear constraints
 - `lavTestWald(fit, constraints = "a == b")`
 - `lavTestWald(fit, constraints = con)`
- Multiple constraints merged in `lavTestWald()`

Testing constraints—examples

34/42

model and constraints

```

1 mod1 <- 'y1 ~ a*x1 + b*x2
2       y2 ~ x1
3       y1~~ c*y1
4       y2~~ d*y2'
5 fit3 <- sem(mod1, data=d)
6 con <- 'a == b
7       c == d^2'
```

Wald test edited output

```

1 > lavTestWald(fit3, constraints = 'a==0')
2 $stat [1] 196.3631; $df [1] 1
3 $p.value [1] 0; $se [1] "standard"
4
5 > lavTestWald(fit3, constraints = con)
6 $stat [1] 16.01391; $df [1] 2
7 $p.value [1] 0.0003331379; $se [1] "standard"
```

Testing constraints—likelihood ratio test

35/42

- Beware, things change if other than ML estimator
- Fit two models: (1) unconstrained, (2) constrained
- Write down log-likelihoods and dfs:

LL_u and df_u

LL_c and df_c

- Compute

$$LR = 2 * (LL_u - LL_c)$$

$$df = df_u - df_c$$

$$p = \text{Prob}(\text{Chi2}(df) > LR)$$

p? R: `pchisq(LR, df, lower.tail = FALSE)`

- Or better use `anova(modu, modc)`

✚ Testing constraints—robust likelihood ratio test

36/42

When estimator is MLM, MLR, or WLSM, the “manual” LR test above is invalid. Use Sattora-Bentler LR test instead:

$c0/c1$: scaling correction factor for M_0 / M_1

$d0/d1$: degrees of freedom for M_0 / M_1

$T0/T1$: Sattora-Bentler scaled χ^2 for M_0 / M_1

$$cd = (d0 * c0 - d1 * c1)/(d0 - d1)$$

$$T = (T0 * c0 - T1 * c1)/cd$$

T is distributed χ^2 with $df = d0 - d1$

The `anova()` function still works...

```
1 fit3 <- sem(mod1, data=d, estimator = 'MLR')  
2 fit2 <- sem(mod2, data=d, estimator = 'MLR')  
3 anova(fit3, fit2)
```

Testing constraints with LR test—Example

37/42

LRtest

```

1 mod2 <- 'y1 ~ a*x1 + b*x2
2       y2 ~ x1 + x2
3       y1~~ c*y1
4       y2~~ d*y2'
5 mod3 <- 'y1 ~ a*x1 + a*x2
6       y2 ~ x1 + x2
7       y1~~ b*y1
8       y2~~ b*y2'
9 fit3 <- sem(mod3, data=d)
10 fit2 <- sem(mod2, data=d)
11 anova(fit3, fit2)

```

LR-test output

```

1 > anova(fit3, fit2)
2 Chi Square Difference Test
3
4      Df      AIC      BIC  Chisq Chisq diff Df diff Pr(>Chisq)
5 fit2  0 1377.1 1395.4  0.000
6 fit3  2 1395.9 1408.9 22.745      22.745      2 1.151e-05 ***
7 ---
8 Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Complicate models using Modification Indices

38/42

- Modification Indices: score tests for parameters currently fixed to 0 or otherwise constrained
- Chi2(1df) distributed if constraint is true
- Score tests are approximate LRtests, guestimated without refitting extended model
- Wald tests are approximate LRtests, guestimated without refitting more restricted model
- check $\text{modindex} \geq 4$ (ex ante) or ≥ 10 (ex post)

```

1 summary(fit3, modindices=TRUE)
2 mi3 <- modindices(fit3)
3 mi3[mi3$op == "~~",]
4 mi3[mi3$mi > 4,]

```

Modification indices (output)

39/42

```

MODINDICES for (Blau-Duncan)
1 summary(fit.bd, modindices = TRUE)
2 [...omitted for brevity...]
3 Modification Indices:
4
5      lhs op   rhs      mi      epc sepc.lv sepc.all sepc.nox
6 7   faed ~~   faed  0.000  0.000   0.000   0.000   0.000
7 8   faed ~~   faocc 0.000  0.000   0.000   0.000   0.000
8 9   faocc ~~   faocc 0.000  0.000   0.000   0.000   0.000
9 12  educ ~~   occ1 13.357 -0.061 -0.061 -0.061 -0.061
10 13  educ ~    occ1 13.357 -0.090 -0.090 -0.090 -0.090
11 14  occ1 ~    faed 13.357  0.025  0.025  0.025  0.025
12 15  faed ~    educ  0.000  0.000  0.000  0.000  0.000
13 16  faed ~    occ1 10.053  0.021  0.021  0.021  0.021
14 17  faed ~   faocc  0.000  0.000  0.000  0.000  0.000
15 18 faocc ~    educ  0.000  0.000  0.000  0.000  0.000
16 19 faocc ~    occ1  4.960 -0.020 -0.020 -0.020 -0.020
17 20 faocc ~    faed  0.000  0.000  0.000  0.000  0.000

```

Note the insensible modifications

Non-recursive path models

40/42

- Causal direct/indirect feedback
- Mutual influence
- Correlated disturbances with direct effects between DVs
- Example: dyadic analysis – friends, siblings, influence each other...
- Identification is not guaranteed

Non-recursive path model

41/42

- Direct feedback loop
- Infinite number of indirect effects

unidentified non-recursive

```
1 mod3 <- 'y1 ~ a*x1 + x2
2         y2 ~ b*x2 + x1
3         y1 ~ c*y2
4         y2 ~ d*y1
5         y1 ~~ y2
6 summary(fit3 <- sem(mod3, data=d))
```

identified non-recursive

```
1 mod4 <- 'y1 ~ a*x1
2         y2 ~ b*x2
3         y1 ~ c*y2
4         y2 ~ d*y1
5         y1 ~~ y2
6         e := (a*c)/(1-(c*d))'
7 summary(fit4 <- sem(mod4, data=d))
```

Identification

42/42

- Identification: existence of unique set of parameter estimates
- Possible: some parameters identified, others unidentified
- Necessary but not sufficient conditions:
 - ① $Df \geq 0$ (N-parameters \geq N-obs)
 - ② Every latent variable (e.g., disturbance, factor) has a scale
- Underidentification ($Df < 0$): multiple parameters lead to same predictions, which ones should be reported?
- Example: EFAs are underidentified, enabling rotation
- Recursive models with (1) and (2) are identified
- For non-recursive models difficult to assess identification
 - order condition: For an equation to be identified, the number of **excluded** exogenous variables for each endogenous variable must be at least as large as the number of **total** endogenous variables, minus one.
 - rank condition: definite answer, but complicated to assess
 - empirical unidentification
 - run model with artificial data before data collection