Structural Equation Modeling with R and lavaan

Day 2: Full Structural Regression Models

Constantin Manuel Bosancianu

WZB Berlin Social Science Center
Institutions and Political Inequality
manuel.bosancianu@gmail.com

September 21, 2021

Today's plan

Less theory than yesterday, to allow for more time in the lab session to explore **lavaan**:

- ✓ basic options for estimation of path models

Today's plan

Less theory than yesterday, to allow for more time in the lab session to explore **lavaan**:

- ✓ basic options for estimation of path models
- ✓ assessing model fit

Today's plan

Less theory than yesterday, to allow for more time in the lab session to explore **lavaan**:

- ✓ basic options for estimation of path models
- ✓ assessing model fit
- ✓ full structural regression models

Today's plan

Less theory than yesterday, to allow for more time in the lab session to explore **lavaan**:

- ✓ basic options for estimation of path models
- ✓ assessing model fit
- ✓ full structural regression models

Best practices in analyzing and presenting results from these models to an audience.

Estimation

Strategies of estimation

Two broad classes for path models:

- ✓ **single-equation methods**: focus on each equation in the system at a time

Strategies of estimation

Two broad classes for path models:

- ✓ **single-equation methods**: focus on each equation in the system at a time
- ✓ **simultaneous methods**: estimate all parameters in one go

Strategies of estimation

Two broad classes for path models:

- ✓ **single-equation methods**: focus on each equation in the system at a time
- ✓ **simultaneous methods**: estimate all parameters in one go

Advantages and disadvantages to each, though the **simultaneous** approach tends to produce estimates with lower variance (is more *efficient*).

Single-equation methods (I)

Estimate one-by-one an equation for each endogenous variable in the *recursive* model using OLS.

Advantages:

- ✓ less sensitive to specification error, as it “compartmentalizes” model in isolated parts

Single-equation methods (I)

Estimate one-by-one an equation for each endogenous variable in the *recursive* model using OLS.

Advantages:

- ✓ less sensitive to specification error, as it "compartmentalizes" model in isolated parts
- ✓ don't require an identified model, or assume multivariate normality

Single-equation methods (I)

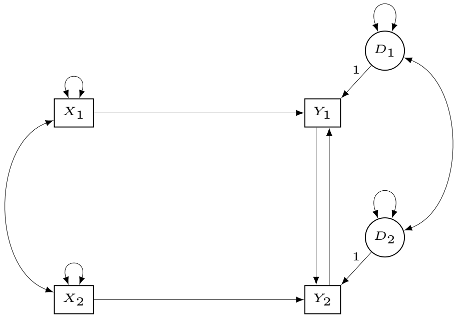
Estimate one-by-one an equation for each endogenous variable in the *recursive* model using OLS.

Advantages:

- ✓ less sensitive to specification error, as it "compartmentalizes" model in isolated parts
- ✓ don't require an identified model, or assume multivariate normality

Main disadvantage: they don't provide a measure of *global* model fit.

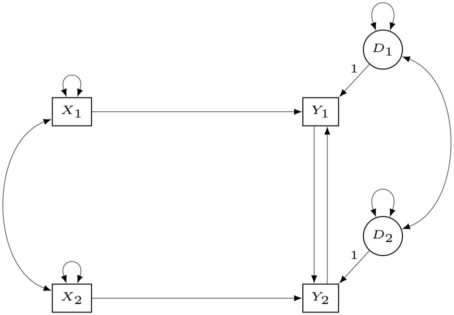
Single-equation methods (II)



Standard regression

Standard OLS no longer works for *non-recursive* specifications (regression residuals are no longer independent of predictors).

Single-equation methods (II)



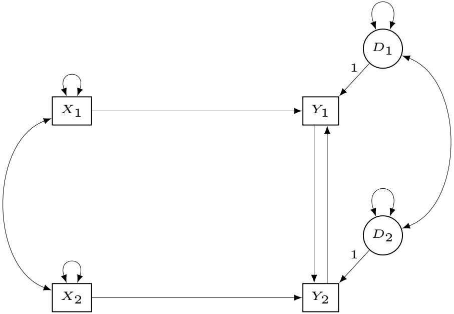
Standard regression

Standard OLS no longer works for *non-recursive* specifications (regression residuals are no longer independent of predictors).

2-stage least squares (2SLS) (something also partly attributed to Sewall Wright!):

- ✓ regress "problematic" variable (Y_1) on instrument (X_1), and save predicted values $\hat{Y}_1$

Single-equation methods (II)



Standard regression

Standard OLS no longer works for *non-recursive* specifications (regression residuals are no longer independent of predictors).

2-stage least squares (2SLS) (something also partly attributed to Sewall Wright!):

- regress "problematic" variable (Y_1) on instrument (X_1), and save predicted values $\hat{Y}_1$
- regress Y_2 on predicted values $\hat{Y}_1$

Simultaneous methods

All free parameters are estimated in one go $\Rightarrow$ the model must be correctly specified.

Simultaneous methods

All free parameters are estimated in one go $\Rightarrow$ the model must be correctly specified.

If not, we get *propagation of specification error*: bias in one parameter estimate affects the other estimates as well.

Simultaneous methods

All free parameters are estimated in one go $\Rightarrow$ the model must be correctly specified.

If not, we get *propagation of specification error*: bias in one parameter estimate affects the other estimates as well.

The most common such method is **maximum likelihood**, which comes in a variety of "flavors"; great advantage in that it works for latent variables as well.

Maximum Likelihood (I)

Widespread usage in statistics. **Principle**: find estimates that maximize the probability of seeing the data (the covariances) in our sample.

Maximum Likelihood (I)

Widespread usage in statistics. **Principle**: find estimates that maximize the probability of seeing the data (the covariances) in our sample.

A *fit function* is minimized in an iterative process, until it produces the smallest difference between predicted covariances (under the model) and observed ones (in the data).

Maximum Likelihood (I)

Widespread usage in statistics. **Principle**: find estimates that maximize the probability of seeing the data (the covariances) in our sample.

A *fit function* is minimized in an iterative process, until it produces the smallest difference between predicted covariances (under the model) and observed ones (in the data).

This is not a simple calculation like with OLS, but a iterative procedure: update coefficients $\Rightarrow$ check difference $\Rightarrow$ update $\Rightarrow$ check again . . .

Maximum Likelihood (I)

Widespread usage in statistics. **Principle**: find estimates that maximize the probability of seeing the data (the covariances) in our sample.

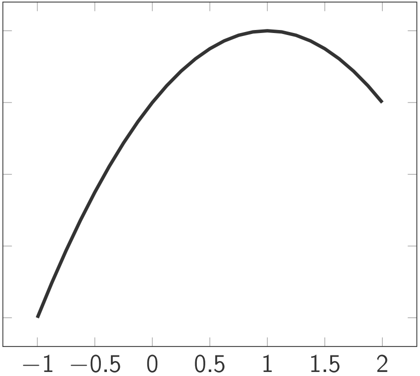
A *fit function* is minimized in an iterative process, until it produces the smallest difference between predicted covariances (under the model) and observed ones (in the data).

This is not a simple calculation like with OLS, but a iterative procedure: update coefficients $\Rightarrow$ check difference $\Rightarrow$ update $\Rightarrow$ check again ...

Problem if there are *multiple optima*: multiple sets of estimates that produce the same degree of fit between model and data.

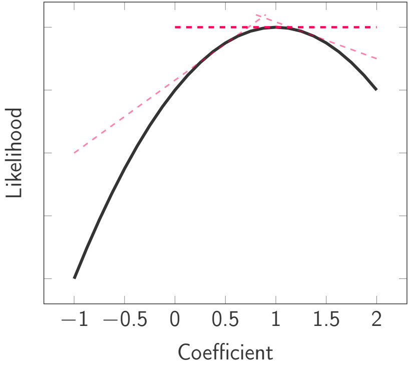
Maximum Likelihood (II)

Convergence of the algorithm is reached when the change in fit is extremely small.



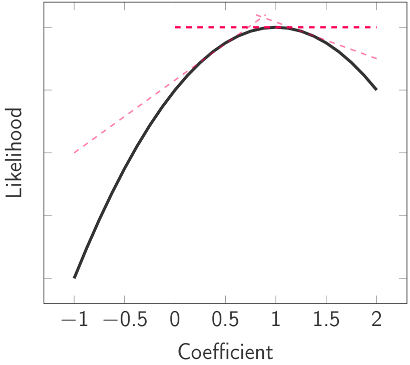
Maximum Likelihood (III)

Setting the first derivative to the likelihood function to 0 gives you the coefficients.



Maximum Likelihood (III)

Setting the first derivative to the likelihood function to 0 gives you the coefficients.



The second derivative to the likelihood function is used to determine if we found the minimum or maximum,[1] as well as to compute the standard errors.

[1] If it's negative, we found the maximum; if it's positive, it's a minimum.

Robust ML

Robust (MLR): for continuous endogenous variables with non-normal distributions. Data is analyzed with standard ML, but SEs and model test statistics are adjusted for non-normality.

Robust ML

Robust (MLR): for continuous endogenous variables with non-normal distributions. Data is analyzed with standard ML, but SEs and model test statistics are adjusted for non-normality.

An alternative is the **Bollen–Stine bootstrap**:

- ✓ default ML is used repeatedly on samples drawn with replacement from the working sample

Robust ML

Robust (MLR): for continuous endogenous variables with non-normal distributions. Data is analyzed with standard ML, but SEs and model test statistics are adjusted for non-normality.

An alternative is the **Bollen–Stine bootstrap**:

- ✓ default ML is used repeatedly on samples drawn with replacement from the working sample
- ✓ βs and SEs are obtained from the empirical sampling distributions of the estimates

Robust ML

Robust (MLR): for continuous endogenous variables with non-normal distributions. Data is analyzed with standard ML, but SEs and model test statistics are adjusted for non-normality.

An alternative is the **Bollen–Stine bootstrap**:

- ✓ default ML is used repeatedly on samples drawn with replacement from the working sample
- ✓ βs and SEs are obtained from the empirical sampling distributions of the estimates
- ✓ the *p* values for estimates and model fit tests are also based on these replications

Categorical outcomes

With at least 6–7 categories, and distributions that approximate a Gaussian one, ML should provide reasonably accurate estimates (Rhemtulla, Brosseau-Liard, & Savalei, 2012).

Categorical outcomes

With at least 6–7 categories, and distributions that approximate a Gaussian one, ML should provide reasonably accurate estimates (Rhemtulla et al., 2012).

Alternatives for variables with fewer categories:

1. *weighted least squares* (WLS): also incorporates variance of observations in estimation
2. *robust WLS*

Categorical outcomes

With at least 6–7 categories, and distributions that approximate a Gaussian one, ML should provide reasonably accurate estimates (Rhemtulla et al., 2012).

Alternatives for variables with fewer categories:

1. *weighted least squares* (WLS): also incorporates variance of observations in estimation
2. *robust WLS*
3. *full-information ML* based on numerical integration—complex in terms of calculations (read: **slow**), but implemented in software

Model fit

Local fit (I)

The goal is to understand for small subsets of a larger model, whether predicted covariances (under the model) fit observed covariances.

If not, where are the biggest gaps, and what can we learn from these patterns?

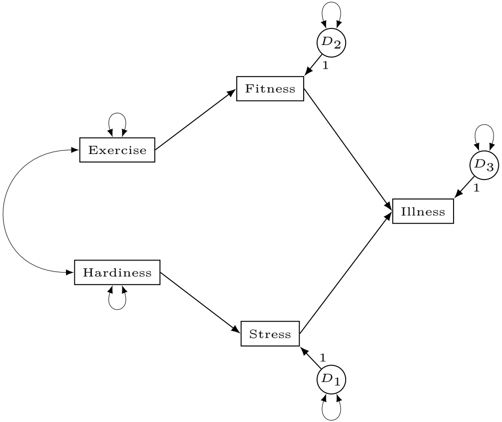
Local fit (I)

The goal is to understand for small subsets of a larger model, whether predicted covariances (under the model) fit observed covariances.

If not, where are the biggest gaps, and what can we learn from these patterns?

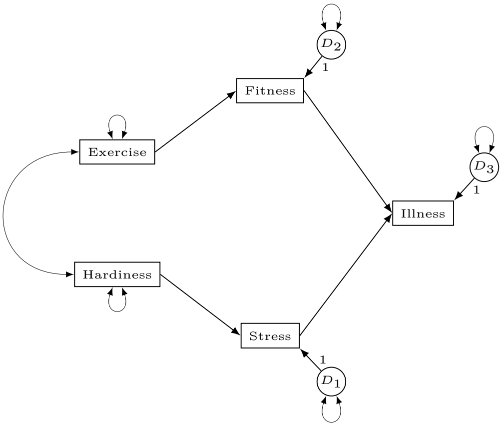
Roth et al. (1989) test an explanatory model of students' susceptibility to illness based on exercise, psychological hardiness, fitness, and stress.

Explaining illness



Exercise and psychological hardiness display no direct effect on illness ⇒ full mediation.

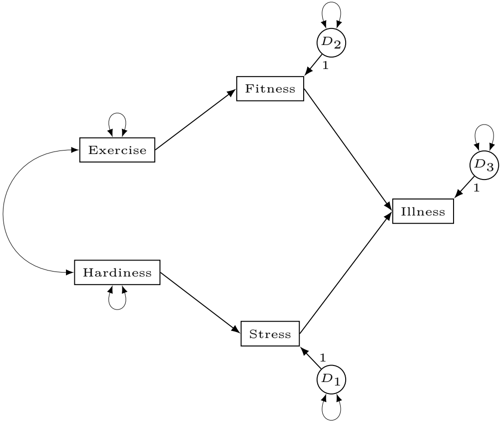
Explaining illness



Exercise and psychological hardiness display no direct effect on illness $\Rightarrow$ full mediation.

The goal is to understand how the multiple components of the model fit the data, under the assumption of a correct model specification.

Explaining illness



Exercise and psychological hardiness display no direct effect on illness $\Rightarrow$ full mediation.

The goal is to understand how the multiple components of the model fit the data, under the assumption of a correct model specification.

Once we control for hardiness, exercise and stress should be independent of each other. 4 more such conditional independences can be evaluated.

Local fit (II)

Any absolute discrepancy between predicted and observed correlations larger than 0.10 is a sign of a poor local fit.

Local fit (II)

Any absolute discrepancy between predicted and observed correlations larger than 0.10 is a sign of a poor local fit.

Independence	Controlling for	Partial correlation
Exercise $\perp\!\!\!\perp$ Stress	Hardiness	-0.058
Exercise $\perp\!\!\!\perp$ Illness	Fitness & Stress	0.039
Hardiness $\perp\!\!\!\perp$ Fitness	Exercise	0.089
Hardiness $\perp\!\!\!\perp$ Illness	Fitness & Stress	-0.081
Fitness $\perp\!\!\!\perp$ Stress	Exercise & Hardiness	-0.103

Local fit (II)

Any absolute discrepancy between predicted and observed correlations larger than 0.10 is a sign of a poor local fit.

Independence	Controlling for	Partial correlation
Exercise $\perp\!\!\!\perp$ Stress	Hardiness	−0.058
Exercise $\perp\!\!\!\perp$ Illness	Fitness & Stress	0.039
Hardiness $\perp\!\!\!\perp$ Fitness	Exercise	0.089
Hardiness $\perp\!\!\!\perp$ Illness	Fitness & Stress	−0.081
Fitness $\perp\!\!\!\perp$ Stress	Exercise & Hardiness	−0.103

It's possible we need to specify a causal link between fitness and stress (though the estimate is at the border).

After fit assessment

Estimation: Kline (2015) has an example for a *single-equation* approach (pp. 241–247), but here we jump directly to the *simultaneous* approach.

After fit assessment

Estimation: Kline (2015) has an example for a *single-equation* approach (pp. 241–247), but here we jump directly to the *simultaneous* approach.

The results presented are based on the default ML algorithm in lavaan, fitted to the covariance matrix of the model.

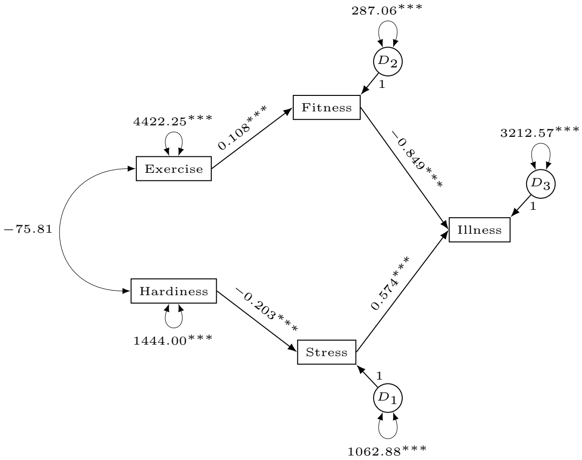
After fit assessment

Estimation: Kline (2015) has an example for a *single-equation* approach (pp. 241–247), but here we jump directly to the *simultaneous* approach.

The results presented are based on the default ML algorithm in lavaan, fitted to the covariance matrix of the model.

This also produces the measures of global fit we discuss in the next subsection, though these are not reported here.

Results simultaneous approach (I)



Unstandardized estimates (SEs not depicted)

Results simultaneous approach (II)

We can also define in lavaan, and estimate, a series of indirect effects:

- ✓ Exercise $\rightarrow$ Illness: -0.092^{***} (0.021)

Results simultaneous approach (II)

We can also define in lavaan, and estimate, a series of indirect effects:

- ✓ Exercise → Illness: -0.092^{***} (0.021)
- ✓ Hardiness → Illness: -0.116^{***} (0.031)

Results simultaneous approach (II)

We can also define in lavaan, and estimate, a series of indirect effects:

- ✓ Exercise → Illness: -0.092^{***} (0.021)
- ✓ Hardiness → Illness: -0.116^{***} (0.031)

If doing this by hand (multiplying direct effects), there is a special approximate formula for the SE of the indirect effect (Sobel, 1982):

$$SE_{ab} = \sqrt{b^2 SE_a^2 + a^2 SE_b^2} \tag{1}$$

Residuals (I)

lavaan also produces **correlation** (or covariance) **residuals**: difference between observed and predicted correlations (covariances).

Residuals (I)

lavaan also produces **correlation** (or covariance) **residuals**: difference between observed and predicted correlations (covariances).

	Fitness	Stress	Illness	Exercise	Hardiness
Fitness	0.000				
Stress	-0.133	0.000			
Illness	-0.038	0.030	0.000		
Exercise	0.000	-0.057	0.016	0.000	
Hardiness	0.082	0.000	-0.091	0.000	0.000

Correlation residuals

Residuals (I)

lavaan also produces **correlation** (or covariance) **residuals**: difference between observed and predicted correlations (covariances).

	Fitness	Stress	Illness	Exercise	Hardiness
Fitness	0.000				
Stress	-0.133	0.000			
Illness	-0.038	0.030	0.000		
Exercise	0.000	-0.057	0.016	0.000	
Hardiness	0.082	0.000	-0.091	0.000	0.000

Correlation residuals

As before, absolute values that are larger than 0.10 are problematic.

Residuals (II)

We can also obtain **standardized residuals**: a ratio or covariance residuals and their SEs.

Residuals (II)

We can also obtain **standardized residuals**: a ratio or covariance residuals and their SEs.

These approximate a z distribution, allowing inference to the population.

Residuals (II)

We can also obtain **standardized residuals**: a ratio or covariance residuals and their SEs.

These approximate a *z* distribution, allowing inference to the population.

	Fitness	Stress	Illness	Exercise	Hardiness
Fitness	0.000				
Stress	-2.548	0.000			
Illness	-2.573	2.573	0.000		
Exercise	0.000	-1.128	0.358	0.000	
Hardiness	1.708	0.000	-1.921	0.000	0.000

Standardized residuals

Global fit indices

Single-value summaries of the fit of the *entire* model specification to the data.

Global fit indices

Single-value summaries of the fit of the *entire* model specification to the data.

We cover 2 families: (1) model test statistics, and (2) approximate fit indices.

Global fit indices

Single-value summaries of the fit of the *entire* model specification to the data.

We cover 2 families: (1) model test statistics, and (2) approximate fit indices.

The usual disclaimers apply as for any single-number summary of fit:

- ✓ even a good value can hide poor fit in parts of the model

Global fit indices

Single-value summaries of the fit of the *entire* model specification to the data.

We cover 2 families: (1) model test statistics, and (2) approximate fit indices.

The usual disclaimers apply as for any single-number summary of fit:

- ✓ even a good value can hide poor fit in parts of the model
- ✓ hard to identify where the problem is in the case of poor fit

Global fit indices

Single-value summaries of the fit of the *entire* model specification to the data.

We cover 2 families: (1) model test statistics, and (2) approximate fit indices.

The usual disclaimers apply as for any single-number summary of fit:

- ✓ even a good value can hide poor fit in parts of the model
- ✓ hard to identify where the problem is in the case of poor fit
- ✓ good fit is not synonymous with theoretical soundness

Model test statistics

Based on the model χ^2 and test the **exact fit**: is the model-predicted covariance matrix considerably *different* than the sample covariance matrix?

Model test statistics

Based on the model χ^2 and test the **exact fit**: is the model-predicted covariance matrix considerably *different* than the sample covariance matrix?

This is an **accept-support test**: we *don't* want to reject the null of no difference in this case!

Model test statistics

Based on the model χ^2 and test the **exact fit**: is the model-predicted covariance matrix considerably *different* than the sample covariance matrix?

This is an **accept-support test**: we *don't* want to reject the null of no difference in this case!

If we do reject H_0 then our model does not do a good job at capturing reality.

Model test statistics

Based on the model χ^2 and test the **exact fit**: is the model-predicted covariance matrix considerably *different* than the sample covariance matrix?

This is an **accept-support test**: we *don't* want to reject the null of no difference in this case!

If we do reject H_0 then our model does not do a good job at capturing reality.

Never enough by itself; most valuable when used in conjunction with local fit testing.

Model χ^2

$$\chi^2 = (N - 1) * F_{ML} \tag{2}$$

N is sample size, and F_{ML} is the fit function minimized in ML estimation.[2]

[2] Other software implements this as $N * F_{ML}$, but asymptotically they're the same thing.

Model χ^2

$$\chi^2 = (N - 1) * F_{ML} \tag{2}$$

N is sample size, and F_{ML} is the fit function minimized in ML estimation.[2]

Follows a χ^2 distribution with degrees of freedom df_M.

[2] Other software implements this as $N * F_{ML}$, but asymptotically they're the same thing.

Model χ^2

$$\chi^2 = (N - 1) * F_{ML} \tag{2}$$

N is sample size, and F_{ML} is the fit function minimized in ML estimation.[2]

Follows a χ^2 distribution with degrees of freedom df_M.

If $\chi^2 = 0$, then $F_{ML} = 0$, so there is no discrepancy between model and sample covariances. Higher values of χ^2 denote worse fit (a "badness-of-fit" indicator).

[2] Other software implements this as $N * F_{ML}$, but asymptotically they're the same thing.

χ^2 sensitivity

Easy to reduce by freeing parameters. Other influences on value:

1. **non-normality** (difficult to predict direction of bias)

χ^2 sensitivity

Easy to reduce by freeing parameters. Other influences on value:

1. **non-normality** (difficult to predict direction of bias)
2. **correlation magnitude**: allows for larger discrepancies between covariance matrices

χ^2 sensitivity

Easy to reduce by freeing parameters. Other influences on value:

1. **non-normality** (difficult to predict direction of bias)
2. **correlation magnitude**: allows for larger discrepancies between covariance matrices
3. **unique variance** in indicators

χ^2 sensitivity

Easy to reduce by freeing parameters. Other influences on value:

1. **non-normality** (difficult to predict direction of bias)
2. **correlation magnitude**: allows for larger discrepancies between covariance matrices
3. **unique variance** in indicators
4. **sample size**: larger samples tend to produce higher χ^2 values

χ^2 sensitivity

Easy to reduce by freeing parameters. Other influences on value:

1. **non-normality** (difficult to predict direction of bias)
2. **correlation magnitude**: allows for larger discrepancies between covariance matrices
3. **unique variance** in indicators
4. **sample size**: larger samples tend to produce higher χ^2 values

χ^2 variants

Various proposals for adjustments to χ^2.

With robust ML (MLR), you can obtain a **Satorra–Bentler scaled** χ^2, which applies a scaling correction factor (the average kurtosis in the data) to the model χ^2.

$$\chi^2_{SB} = \frac{\chi^2_M}{c} \tag{3}$$

χ^2 variants

Various proposals for adjustments to χ^2.

With robust ML (MLR), you can obtain a **Satorra–Bentler scaled** χ^2, which applies a scaling correction factor (the average kurtosis in the data) to the model χ^2.

$$\chi^2_{SB} = \frac{\chi^2_M}{c} \tag{3}$$

There is also a **Satorra–Bentler adjusted** χ^2, though it's less used than the scaled version.

Approximate fit indices

They do not rely on a significance test, but rather offer a numerical continuous summary of the model–data fit. Typically come standardized between 0–1.

Approximate fit indices

They do not rely on a significance test, but rather offer a numerical continuous summary of the model–data fit. Typically come standardized between 0–1.

Types:

- ✓ **absolute fit**: comparison of model-implied covariances and data covariances

Approximate fit indices

They do not rely on a significance test, but rather offer a numerical continuous summary of the model–data fit. Typically come standardized between 0–1.

Types:

- **absolute fit**: comparison of model-implied covariances and data covariances
- **incremental (relative, comparative) fit**: relative improvement over a baseline model

Approximate fit indices

They do not rely on a significance test, but rather offer a numerical continuous summary of the model–data fit. Typically come standardized between 0–1.

Types:

- **absolute fit**: comparison of model-implied covariances and data covariances
- **incremental (relative, comparative) fit**: relative improvement over a baseline model
- **parsimony-adjusted**: incorporate a penalty for model complexity

Approximate fit indices

They do not rely on a significance test, but rather offer a numerical continuous summary of the model–data fit. Typically come standardized between 0–1.

Types:

- **absolute fit**: comparison of model-implied covariances and data covariances
- **incremental (relative, comparative) fit**: relative improvement over a baseline model
- **parsimony-adjusted**: incorporate a penalty for model complexity
- **predictive fit**: hypothetical fit in other samples from same population

RMSEA (I)

RMSEA = root mean squared error of approximation ($\hat{\varepsilon}$).

An absolute fit index that measures departure from close fit (it's a "badness-of-fit" indicator).

RMSEA (I)

RMSEA = root mean squared error of approximation ($\hat{\varepsilon}$).

An absolute fit index that measures departure from close fit (it's a "badness-of-fit" indicator).

This limit of close fit is defined as

$$\hat{\Delta}_M = max(0, \chi^2_M - df_M) \tag{4}$$

RMSEA (I)

RMSEA = root mean squared error of approximation ($\hat{\varepsilon}$).

An absolute fit index that measures departure from close fit (it's a "badness-of-fit" indicator).

This limit of close fit is defined as

$$\hat{\Delta}_M = max(0, \chi^2_M - df_M) \tag{4}$$

If $\hat{\Delta}_M = 0$, there is no departure from close fit. Otherwise:

$$\hat{\varepsilon} = \sqrt{\frac{\hat{\Delta}_M}{df_M(N-1)}} \tag{5}$$

RMSEA (II)

Typically, also gets reported with a 90% CI: $[\hat{\varepsilon}_L, \hat{\varepsilon}_U]$.

RMSEA (II)

Typically, also gets reported with a 90% CI: $[\hat{\varepsilon}_L, \hat{\varepsilon}_U]$.

Hard to put thresholds on this, but if the upper bound of CI intersects 0.10 caution is warranted.

RMSEA (II)

Typically, also gets reported with a 90% CI: $[\hat{\varepsilon}_L, \hat{\varepsilon}_U]$.

Hard to put thresholds on this, but if the upper bound of CI intersects 0.10 caution is warranted.

Using the Satorra–Bentler scaled χ^2 produces a more robust version of the RMSEA to departures from normality.

SRMR

Standardized root mean squared residual: an absolute fit index where higher values denote poorer fit.

SRMR

Standardized root mean squared residual: an absolute fit index where higher values denote poorer fit.

Computed as the squared root of the average squared covariance residual (in standardized format) $\Rightarrow$ mean absolute correlation residual.

SRMR

Standardized root mean squared residual: an absolute fit index where higher values denote poorer fit.

Computed as the squared root of the average squared covariance residual (in standardized format) $\Rightarrow$ mean absolute correlation residual.

Values greater than 0.10 indicate poor fit.

CFI

The Bentler **comparative fit index** (CFI) is an incremental (relative) fit index: compares the proposed model to an independence (null) model.

CFI

The Bentler **comparative fit index** (CFI) is an incremental (relative) fit index: compares the proposed model to an independence (null) model.

It's therefore a "goodness-of-fit" measure.

$$CFI = 1 - \frac{\hat{\Delta}_M}{\hat{\Delta}_B} \tag{6}$$

where for the baseline model $\hat{\Delta}_B = max(0, \chi^2_B - df_B)$.

CFI

The Bentler **comparative fit index** (CFI) is an incremental (relative) fit index: compares the proposed model to an independence (null) model.

It's therefore a "goodness-of-fit" measure.

$$CFI = 1 - \frac{\hat{\Delta}_M}{\hat{\Delta}_B} \tag{6}$$

where for the baseline model $\hat{\Delta}_B = max(0, \chi^2_B - df_B)$.

CFI of 0.90 would indicate model fit that is about 90% better than the baseline model.

Usage recommendations

Report the model χ^2 and its degrees of freedom and p value. Irrespective of whether this test is passed or not, conduct local fit testing.

Usage recommendations

Report the model χ^2 and its degrees of freedom and p value. Irrespective of whether this test is passed or not, conduct local fit testing.

Report the matrix of correlation residuals, even if only in the appendix.

Usage recommendations

Report the model χ^2 and its degrees of freedom and *p* value. Irrespective of whether this test is passed or not, conduct local fit testing.

Report the matrix of correlation residuals, even if only in the appendix.

Report values from approximate fit indices as well (RMSEA, CFI, SRMR), but keep in mind that thresholds for them depend considerably on (1) sample size, (2) distributional assumptions, (3) degree of mis-specification, and (4) estimation method (Xia & Yang, 2019).

Nested models

2 models are nested if one includes all the parameters of the other, plus at least one more. Example: one parameter is freed or constrained.

Nested models

2 models are nested if one includes all the parameters of the other, plus at least one more. Example: one parameter is freed or constrained.

Variants:

- ✓ **model trimming**: start with a complex model and constrain parameters to 0—ideally, χ^2_M doesn't increase much

Nested models

2 models are nested if one includes all the parameters of the other, plus at least one more. Example: one parameter is freed or constrained.

Variants:

- **model trimming**: start with a complex model and constrain parameters to 0—ideally, χ^2_M doesn't increase much
- **model building**: start with a simple model and free parameters—ideally, χ^2_M decreases considerably

Nested models

2 models are nested if one includes all the parameters of the other, plus at least one more. Example: one parameter is freed or constrained.

Variants:

- **model trimming**: start with a complex model and constrain parameters to 0—ideally, χ^2_M doesn't increase much
- **model building**: start with a simple model and free parameters—ideally, χ^2_M decreases considerably

Can be compared with the χ^2 difference statistic, with *df* the difference in the number of parameters between the two models.

Non-nested models (I)

Not all models are nested: as an example, freeing some parameters and constraining others in model B compared to A.

Non-nested models (I)

Not all models are nested: as an example, freeing some parameters and constraining others in model B compared to A.

In these instances, the χ^2 difference statistic no longer applies.

Non-nested models (I)

Not all models are nested: as an example, freeing some parameters and constraining others in model B compared to A.

In these instances, the χ^2 difference statistic no longer applies.

Akaike Information Criterion (AIC) allows for such comparisons:

$$AIC = \chi^2_M + 2k \tag{7}$$

where k is the number of estimated (free) parameters.

Non-nested models (I)

Not all models are nested: as an example, freeing some parameters and constraining others in model B compared to A.

In these instances, the χ^2 difference statistic no longer applies.

Akaike Information Criterion (AIC) allows for such comparisons:

$$AIC = \chi^2_M + 2k \tag{7}$$

where k is the number of estimated (free) parameters.

Models need to be estimated on the same sample to compare their AICs.

Non-nested models (II)

The Bayesian Information Criterion (BIC), however, does incorporate sample size differences:

$$BIC = \chi^2_M + k * ln(N) \tag{8}$$

where k is the number of free parameters and N is the sample size.

Non-nested models (II)

The Bayesian Information Criterion (BIC), however, does incorporate sample size differences:

$$BIC = \chi^2_M + k * ln(N) \tag{8}$$

where k is the number of free parameters and N is the sample size.

BIC should be used carefully, as with increasing sample size χ^2_M also may increase even though model fit wouldn't change.

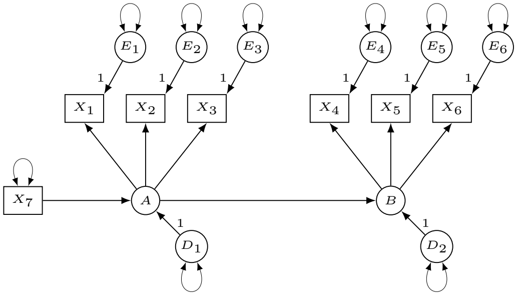
Structural Regression Models

SR models (I)

A **structural regression** model combines a *measurement* part with a *structural* one.

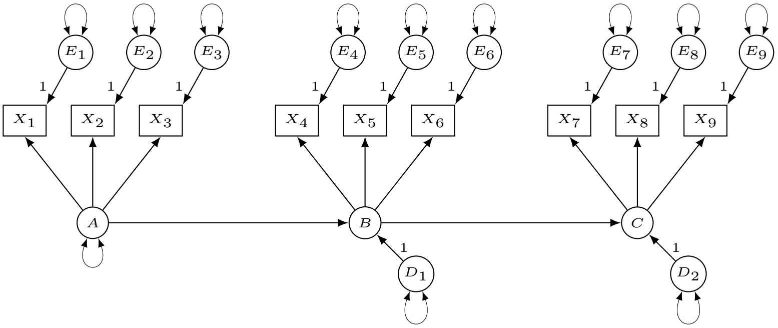
SR models (I)

A **structural regression** model combines a *measurement* part with a *structural* one.



Partially latent SR model

SR models (II)



Fully latent SR model

Identification

Identification is done separately for the measurement and structural part of the model.

Identification

Identification is done separately for the measurement and structural part of the model.

The same principles from yesterday's discussion of path models and factor analysis models apply here.

Identification

Identification is done separately for the measurement and structural part of the model.

The same principles from yesterday's discussion of path models and factor analysis models apply here.

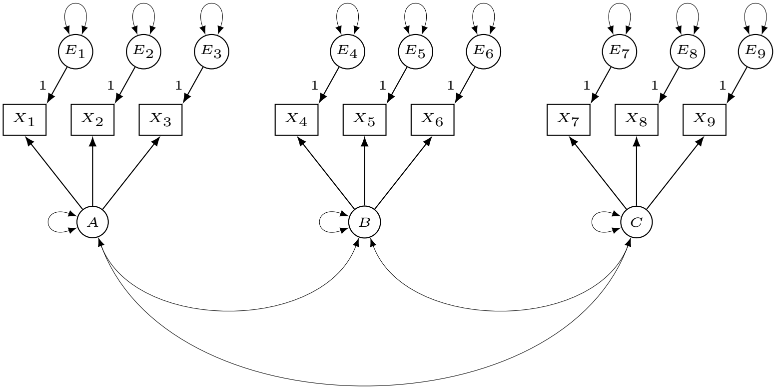
Both parts have to be identified (**two-step identification rule**).

2-step modeling (I)

First, specify a fully latent SR model as a CFA measurement model, to determine whether it fits the data.

2-step modeling (I)

First, specify a fully latent SR model as a CFA measurement model, to determine whether it fits the data.



2-step modeling (II)

Second, convert covariances in the CFA measurement model into causal paths between latents, and check model fit changes.

2-step modeling (II)

Second, convert covariances in the CFA measurement model into causal paths between latents, and check model fit changes.

The hope is that the switch from 1st to 2nd stage does not impact factor loading considerably.

2-step modeling (II)

Second, convert covariances in the CFA measurement model into causal paths between latents, and check model fit changes.

The hope is that the switch from 1st to 2nd stage does not impact factor loading considerably.

A 4-step approach was also proposed, where 1st stage is an EFA measurement model, which can then be reduced to a CFA one in the 2nd stage.

Best practices

A few resources

All of these are currently uploaded on Moodle:

1. Hoyle and Isherwood (2013) offer a checklist of points that should be discussed when reporting SEM results (so do Mueller & Hancock, 2008)

A few resources

All of these are currently uploaded on Moodle:

1. Hoyle and Isherwood (2013) offer a checklist of points that should be discussed when reporting SEM results (so do Mueller & Hancock, 2008)
2. Jackson et al. (2009) offer reporting recommendations for CFA analyses

A few resources

All of these are currently uploaded on Moodle:

1. Hoyle and Isherwood (2013) offer a checklist of points that should be discussed when reporting SEM results (so do Mueller & Hancock, 2008)
2. Jackson et al. (2009) offer reporting recommendations for CFA analyses
3. Mueller and Hancock (2019) provide a handy list of recommendations for reviewers of manuscripts using SEM

A few resources

All of these are currently uploaded on Moodle:

1. Hoyle and Isherwood (2013) offer a checklist of points that should be discussed when reporting SEM results (so do Mueller & Hancock, 2008)
2. Jackson et al. (2009) offer reporting recommendations for CFA analyses
3. Mueller and Hancock (2019) provide a handy list of recommendations for reviewers of manuscripts using SEM
4. Schumacker and Lomax (2016, ch. 16) provide recommendations for modeling, as well as yet another checklist

A few resources

All of these are currently uploaded on Moodle:

1. Hoyle and Isherwood (2013) offer a checklist of points that should be discussed when reporting SEM results (so do Mueller & Hancock, 2008)
2. Jackson et al. (2009) offer reporting recommendations for CFA analyses
3. Mueller and Hancock (2019) provide a handy list of recommendations for reviewers of manuscripts using SEM
4. Schumacker and Lomax (2016, ch. 16) provide recommendations for modeling, as well as yet another checklist
5. Thompson (2000) lists the "10 commandments" of SEM

Selected suggestions: specification

Try to have at least 3 indicators for each factor (even 2 can lead to problems of estimation).

Selected suggestions: specification

Try to have at least 3 indicators for each factor (even 2 can lead to problems of estimation).

Don't specify reciprocal causation unless theory really suggests this is the case.

Selected suggestions: specification

Try to have at least 3 indicators for each factor (even 2 can lead to problems of estimation).

Don't specify reciprocal causation unless theory really suggests this is the case.

Keep in mind the importance of **parsimony**.

Selected suggestions: specification

Try to have at least 3 indicators for each factor (even 2 can lead to problems of estimation).

Don't specify reciprocal causation unless theory really suggests this is the case.

Keep in mind the importance of **parsimony**.

It's fine to constrain parameters, like forcing equality or proportionality, but make sure to defend this theoretically.

Selected suggestions: estimation

Don't rely on standard ML estimation for ordinal variables (6–7); go for full-information ML, or one of the least squares-based methods.

Selected suggestions: estimation

Don't rely on standard ML estimation for ordinal variables (6–7); go for full-information ML, or one of the least squares-based methods.

Don't choose the best model based only on global fit tests and standard thresholds; rather, use local fit assessment—examine correlation residuals' matrices.

Selected suggestions: estimation

Don't rely on standard ML estimation for ordinal variables (6–7); go for full-information ML, or one of the least squares-based methods.

Don't choose the best model based only on global fit tests and standard thresholds; rather, use local fit assessment—examine correlation residuals' matrices.

For a fully latent SR model, don't do "single-shot" estimation. Instead:

1. ensure the measurement model is correctly specified

Selected suggestions: estimation

Don't rely on standard ML estimation for ordinal variables (6–7); go for full-information ML, or one of the least squares-based methods.

Don't choose the best model based only on global fit tests and standard thresholds; rather, use local fit assessment—examine correlation residuals' matrices.

For a fully latent SR model, don't do "single-shot" estimation. Instead:

1. ensure the measurement model is correctly specified
2. after this, proceed to adding causal links between latents

Selected suggestions: others

Respecification is a common part of the modeling process, but it should include theoretical justification.

Selected suggestions: others

Respecification is a common part of the modeling process, but it should include theoretical justification.

If reporting on mediation effects, decompose the total effect into direct and indirect ones, with accompanying SEs and significance tests.

Selected suggestions: others

Respecification is a common part of the modeling process, but it should include theoretical justification.

If reporting on mediation effects, decompose the total effect into direct and indirect ones, with accompanying SEs and significance tests.

Report the residual matrix (at a minimum in the appendix), and discuss it in the main text.

Selected suggestions: others

Respecification is a common part of the modeling process, but it should include theoretical justification.

If reporting on mediation effects, decompose the total effect into direct and indirect ones, with accompanying SEs and significance tests.

Report the residual matrix (at a minimum in the appendix), and discuss it in the main text.

Report measures of model fit in the main text: model χ^2 with *df* and *p* value, RMSEA with 90% CI, CFI (or TLI), and the SRMR.

Thank you for the kind attention!

References I

Hoyle, R. H., & Isherwood, J. C. (2013). Reporting results from structural equation modeling analyses in *Archives of Scientific Psychology*. *Archives of Scientific Psychology*, *1*(1), 14–22.

Jackson, D. L., Gillaspy, J. A., & Purc-Stephenson, R. (2009). Reporting Practices in Confirmatory Factor Analysis: An Overview and Some Recommendations. *Psychological Methods*, *14*(1), 6–23.

Kline, R. B. (2015). *Principles and Practice of Structural Equation Modeling* (4th ed.). New York: The Guilford Press.

Mueller, R. O., & Hancock, G. R. (2008). Best Practices in Structural Equation Modeling. In J. W. Osborne (Ed.), *Best Practices in Quantitative Methods* (pp. 488–508). Thousand Oaks, CA: Sage Publications.

Mueller, R. O., & Hancock, G. R. (2019). Structural Equation Modeling. In G. R. Hancock, L. M. Stapleton, & R. O. Mueller (Eds.), *The Reviewer's Guide to Quantitative Methods in the Social Sciences* (2nd ed., pp. 445–456). New York: Routledge.

References II

Rhemtulla, M., Brosseau-Liard, P. E., & Savalei, V. (2012). When Can Categorical Variables Be Treated as Continuous? A Comparison of Robust Continuous and Categorical SEM Estimation Methods under Suboptimal Conditions. *Psychological Methods*, *17*(3), 354–373.

Roth, D. L., Wiebe, D. J., Fillingim, R. B., & Shay, K. A. (1989). Life Events, Fitness, Hardiness, and Health: A Simultaneous Analysis of Proposed Stress-Resistance Effects. *Journal of Personality and Social Psychology*, *57*(1), 136–142.

Schumacker, R. E., & Lomax, R. G. (2016). *A Beginner's Guide to Structural Equation Modeling* (4th ed.). New York: Routledge.

Sobel, M. E. (1982). Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, *13*, 290–312.

Thompson, B. (2000). Ten Commandments of Structural Equation Modeling. In L. G. Grimm & P. R. Yarnold (Eds.), *Reading and Understanding More Multivariate Statistics* (1st ed., pp. 261–283). Washington, DC: American Psychological Association.

Xia, Y., & Yang, Y. (2019). RMSEA, CFI, and TLI in structural equation modeling with ordered categorical data: The story they tell depends on the estimation methods. *Behavior Research Methods*, *51*(1), 409–428.