Confirmatory Factor Analysis Part Two[1]

STA2053 Fall 2022

[1] See last slide for copyright information.

Original and Surrogate Models

- Original model has expected values, intercepts, and slopes that need not equal one.
- Re-parameterization via a change of variables yields a surrogate model.
- Centered surrogate model has the same covariance matrix as the original.

Why should the variance of the factors equal one?

- Inherited from exploratory factor analysis, which was mostly a disaster.
- The standard answer is something like this: "Because its arbitrary. The variance depends upon the scale on which the variable is measured, but we can't see it to measure it directly. So set it to one for convenience."
- But saying it does not make it so. If F is a random variable with an unknown variance, then
- $Var(F) = \phi$ is an unknown parameter.

Centered Model

One factor, four observed variables

$$
\begin{array}{rcl}
d_1 & = & \lambda_1 F + e_1 \\
d_2 & = & \lambda_2 F + e_2 \\
d_3 & = & \lambda_3 F + e_3 \\
d_4 & = & \lambda_4 F + e_4
\end{array}
$$

$e_1, \ldots, e_4, F$ all independent
$Var(e_j) = \omega_j \quad Var(F) = \phi$
$\lambda_1, \lambda_2, \lambda_3 \neq 0$

$$
\boldsymbol{\Sigma} = \left(\begin{array}{rrrr}
\lambda_1^2\phi + \omega_1 & \lambda_1\lambda_2\phi & \lambda_1\lambda_3\phi & \lambda_1\lambda_4\phi \\
\lambda_1\lambda_2\phi & \lambda_2^2\phi + \omega_2 & \lambda_2\lambda_3\phi & \lambda_2\lambda_4\phi \\
\lambda_1\lambda_3\phi & \lambda_2\lambda_3\phi & \lambda_3^2\phi + \omega_3 & \lambda_3\lambda_4\phi \\
\lambda_1\lambda_4\phi & \lambda_2\lambda_4\phi & \lambda_3\lambda_4\phi & \lambda_4^2\phi + \omega_4
\end{array} \right)
$$

Passes the Counting Rule test with 10 equations in 9 unknowns

But for any $c \neq 0$

$\boldsymbol{\theta}_1$	ϕ	λ_1	λ_2	λ_3	λ_4	ω_1	ω_2	ω_3	ω_4
$\boldsymbol{\theta}_2$	ϕ/c^2	$c\lambda_1$	$c\lambda_2$	$c\lambda_3$	$c\lambda_4$	ω_1	ω_2	ω_3	ω_4

Both yield

$$\boldsymbol{\Sigma} = \begin{pmatrix} \lambda_1^2\phi + \omega_1 & \lambda_1\lambda_2\phi & \lambda_1\lambda_3\phi & \lambda_1\lambda_4\phi \\ \lambda_1\lambda_2\phi & \lambda_2^2\phi + \omega_2 & \lambda_2\lambda_3\phi & \lambda_2\lambda_4\phi \\ \lambda_1\lambda_3\phi & \lambda_2\lambda_3\phi & \lambda_3^2\phi + \omega_3 & \lambda_3\lambda_4\phi \\ \lambda_1\lambda_4\phi & \lambda_2\lambda_4\phi & \lambda_3\lambda_4\phi & \lambda_4^2\phi + \omega_4 \end{pmatrix}$$

The choice $\phi = 1$ just sets $c = \sqrt{\phi}$: convenient but seemingly arbitrary.

Lack of identifiability

- For any set of true parameter values, there are infinitely many untrue sets of parameter values that yield the same $\boldsymbol{\Sigma}$ and hence the same probability distribution of the observable data (assuming multivariate normality).
- There is no way to know the full truth based on the data, no matter how large the sample size.
- But there is a way to know the partial truth.

Certain *functions* of the parameter vector are identifiable

At points in the parameter space where $\lambda_1, \lambda_2, \lambda_3 \neq 0$,

- $\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} = \frac{\lambda_1\lambda_2\phi\lambda_1\lambda_3\phi}{\lambda_2\lambda_3\phi} = \lambda_1^2\phi$
- And so if $\lambda_1 > 0$, the function $\lambda_j\phi^{1/2}$ is identifiable for $j = 1, \ldots, 4$.
- $\sigma_{11} - \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} = \omega_1$, and so ω_j is identifiable for $j = 1, \ldots, 4$.
- $\frac{\sigma_{13}}{\sigma_{23}} = \frac{\lambda_1\lambda_3\phi}{\lambda_2\lambda_3\phi} = \frac{\lambda_1}{\lambda_2}$, so *ratios* of factor loadings are identifiable.

Reliability

- Reliability is the squared correlation between the observed score and the true score.
- The proportion of variance in the observed score that is not error.
- For $D_1 = \lambda_1 F + e_1$ it's

$$
\begin{aligned}
\rho^2 &= \left(\frac{Cov(D_1, F)}{SD(D_1)SD(F)} \right)^2 \\
&= \left(\frac{\lambda_1 \phi}{\sqrt{\lambda_1^2 \phi + \omega_1}\sqrt{\phi}} \right)^2 \\
&= \frac{\lambda_1^2 \phi}{\lambda_1^2 \phi + \omega_1}
\end{aligned}
$$

$\lambda_1^2 \phi$ and ω_1 are both identifiable, so we've got it.

For completeness

$$\rho^2 = \frac{\lambda_1^2\phi}{\lambda_1^2\phi + \omega_1} \qquad \boldsymbol{\Sigma} = \left(\begin{array}{rrrr} \lambda_1^2\phi + \omega_1 & \lambda_1\lambda_2\phi & \lambda_1\lambda_3\phi & \lambda_1\lambda_4\phi \\ \lambda_1\lambda_2\phi & \lambda_2^2\phi + \omega_2 & \lambda_2\lambda_3\phi & \lambda_2\lambda_4\phi \\ \lambda_1\lambda_3\phi & \lambda_2\lambda_3\phi & \lambda_3^2\phi + \omega_3 & \lambda_3\lambda_4\phi \\ \lambda_1\lambda_4\phi & \lambda_2\lambda_4\phi & \lambda_3\lambda_4\phi & \lambda_4^2\phi + \omega_4 \end{array} \right)$$

$$\begin{aligned} \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}\sigma_{11}} &= \frac{\lambda_1\lambda_2\phi\lambda_1\lambda_3\phi}{\lambda_2\lambda_3\phi(\lambda_1^2\phi + \omega_1)} \\ &= \frac{\lambda_1^2\phi}{\lambda_1^2\phi + \omega_1} \\ &= \rho^2 \end{aligned}$$

What can we successfully estimate?

- Error variances are knowable.
- Factor loadings and variance of the factor are not knowable separately.
- But both are knowable up to multiplication by a non-zero constant, so signs of factor loadings are knowable (if one sign is known).
- Relative magnitudes (ratios) of factor loadings are knowable.
- Reliabilities are knowable.

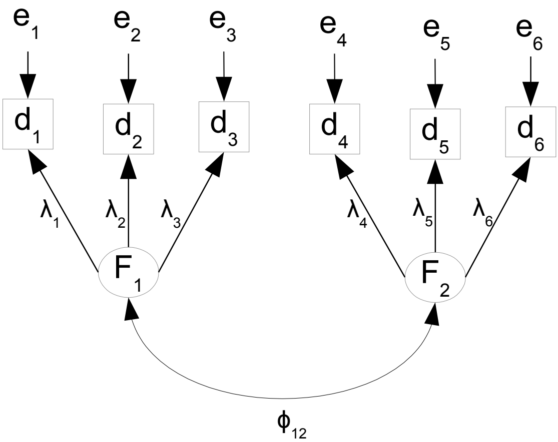
Re-parameterization

- The choice $\phi = 1$ is a very smart re-parameterization.
- It re-expresses the factor loadings as multiples of the square root of ϕ.
- That is, in standard deviation units.
- It preserves what information is accessible about the parameters of the original model.
- Much better than exploratory factor analysis, which lost even the signs of the factor loadings.
- This is the second major re-parameterization. The first was losing the the means and intercepts.

Re-parameterizations

Original model $\rightarrow$ Surrogate model 1 $\rightarrow$ Surrogate model 2 $\ldots$

Add a factor to the centered original model



Model Equations

$$
\begin{aligned}
d_1 &= \lambda_1 F_1 + e_1 \\
d_2 &= \lambda_2 F_1 + e_2 \\
d_3 &= \lambda_3 F_1 + e_3 \\
d_4 &= \lambda_4 F_2 + e_4 \\
d_5 &= \lambda_5 F_2 + e_5 \\
d_6 &= \lambda_6 F_2 + e_6
\end{aligned}
$$

$$cov \begin{pmatrix} F_1 \\ F_2 \end{pmatrix} = \begin{pmatrix} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{pmatrix}$$

$e_1, \ldots, e_6$ independent of each other and of F_1, F_2
$\lambda_1, \ldots \lambda_6 \neq 0$
$Var(e_j) = \omega_j$

Parameters are not identifiable

$\boldsymbol{\Sigma} =$

$$\left(\begin{array}{rrrrrr}
\lambda_1^2\phi_{11}+\omega_1 & \lambda_1\lambda_2\phi_{11} & \lambda_1\lambda_3\phi_{11} & \lambda_1\lambda_4\phi_{12} & \lambda_1\lambda_5\phi_{12} & \lambda_1\lambda_6\phi_{12} \\
\lambda_1\lambda_2\phi_{11} & \lambda_2^2\phi_{11}+\omega_2 & \lambda_2\lambda_3\phi_{11} & \lambda_2\lambda_4\phi_{12} & \lambda_2\lambda_5\phi_{12} & \lambda_2\lambda_6\phi_{12} \\
\lambda_1\lambda_3\phi_{11} & \lambda_2\lambda_3\phi_{11} & \lambda_3^2\phi_{11}+\omega_3 & \lambda_3\lambda_4\phi_{12} & \lambda_3\lambda_5\phi_{12} & \lambda_3\lambda_6\phi_{12} \\
\lambda_1\lambda_4\phi_{12} & \lambda_2\lambda_4\phi_{12} & \lambda_3\lambda_4\phi_{12} & \lambda_4^2\phi_{22}+\omega_4 & \lambda_4\lambda_5\phi_{22} & \lambda_4\lambda_6\phi_{22} \\
\lambda_1\lambda_5\phi_{12} & \lambda_2\lambda_5\phi_{12} & \lambda_3\lambda_5\phi_{12} & \lambda_4\lambda_5\phi_{22} & \lambda_5^2\phi_{22}+\omega_5 & \lambda_5\lambda_6\phi_{22} \\
\lambda_1\lambda_6\phi_{12} & \lambda_2\lambda_6\phi_{12} & \lambda_3\lambda_6\phi_{12} & \lambda_4\lambda_6\phi_{22} & \lambda_5\lambda_6\phi_{22} & \lambda_6^2\phi_{22}+\omega_6
\end{array}\right)$$

$$\begin{array}{rcl}
\boldsymbol{\theta}_1 & = & (\lambda_1, \ldots, \lambda_6, \phi_{11}, \phi_{12}, \phi_{22}, \omega_1, \ldots, \omega_6) \\
\boldsymbol{\theta}_2 & = & (\lambda_1', \ldots, \lambda_6', \phi_{11}', \phi_{12}', \phi_{22}', \omega_1', \ldots, \omega_6')
\end{array}$$

$\lambda_1' = c_1\lambda_1 \quad \lambda_2' = c_1\lambda_2 \quad \lambda_3' = c_1\lambda_3 \quad \phi_{11}' = \phi_{11}/c_1^2$
$\lambda_4' = c_2\lambda_4 \quad \lambda_5' = c_2\lambda_5 \quad \lambda_6' = c_2\lambda_6 \quad \phi_{22}' = \phi_{22}/c_2^2$
$\phi_{12}' = \frac{\phi_{12}}{c_1 c_2}$
$\omega_j' = \omega_j$ for $j = 1, \ldots, 6$

Variances and covariances of factors

- Are knowable only up to multiplication by positive constants.
- Since the parameters of the latent variable model will be recovered from $\mathbf{\Phi} = cov(\mathbf{F})$, they also will be knowable only up to multiplication by positive constants – at best.
- Luckily, in most applications the interest is in testing (pos-neg-zero) more than estimation.

$cov(F_1, F_2)$ is un-knowable, but

- Easy to tell if its zero.
- Sign is known if one factor loading from each set is known – say $\lambda_1 > 0$, $\lambda_4 > 0$.
- And,

$$
\begin{aligned}
\frac{\sigma_{14}}{\sqrt{\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}}\sqrt{\frac{\sigma_{45}\sigma_{46}}{\sigma_{56}}}} &= \frac{\lambda_1\lambda_4\phi_{12}}{\lambda_1\sqrt{\phi_{11}}\lambda_4\sqrt{\phi_{22}}} \\
&= \frac{\phi_{12}}{\sqrt{\phi_{11}}\sqrt{\phi_{22}}} \\
&= Corr(F_1, F_2)
\end{aligned}
$$

- The *correlation* between factors is identifiable!

The correlation between factors is identifiable

- Furthermore, it is the same function of $\mathbf{\Sigma}$ that yields ϕ_{12} under the surrogate model with $Var(F_1) = Var(F_2) = 1$.
- Therefore, $Corr(F_1, F_2) = \phi_{12}$ under the surrogate model is actually $Corr(F_1, F_2)$ under the original model.
- Estimation is very meaningful.

Setting variances of factors to one

- Is a *very* smart re-parameterization.
- Is excellent when the interest is in correlations between factors.

Re-parameterization as a change of variables

$$
\begin{aligned}
d_j &= \lambda_j F_j + e_j \\
&= (\lambda_j \sqrt{\phi_{jj}}) \left(\frac{1}{\sqrt{\phi_{jj}}} F_j \right) + e_j \\
&= \lambda_j' F_j' + e_j
\end{aligned}
$$

Covariances

$$
\begin{aligned}
Cov(F'_j, F'_k) &= Cov\left(\frac{1}{\sqrt{\phi_{jj}}}F_j, \frac{1}{\sqrt{\phi_{kk}}}F_k\right) \\
&= \frac{Cov(F_j, F_k)}{\sqrt{\phi_{jj}}\sqrt{\phi_{kk}}} \\
&= \frac{\phi_{jk}}{\sqrt{\phi_{jj}}\sqrt{\phi_{kk}}} \\
&= Corr(F_j, F_k)
\end{aligned}
$$

Cascading effects

- Understand the re-parameterization as a change of variables.
- Not just an arbitrary restriction of the parameter space.
- It shows there are widespread effects throughout the model.
- Especially if there is a detailed latent variable model.
- Also shows *how* the meanings of other model parameters are affected.

The other standard trick

- Setting variances of all the factors to one is an excellent re-parameterization in disguise.
- The other standard trick is to set a factor loading equal to one for each factor.
- $d = F + e$ is hard to believe if you take it literally.
- It's actually a re-parameterization.
- Every model you've seen with a factor loading of one is a surrogate model.

Back to a single-factor model with $\lambda_1 > 0$

$$\begin{aligned} d_1 &= \lambda_1 F + e_1 \\ d_2 &= \lambda_2 F + e_2 \\ d_3 &= \lambda_3 F + e_3 \\ &\vdots \end{aligned}$$

$$\begin{aligned} d_j &= \left(\frac{\lambda_j}{\lambda_1}\right)(\lambda_1 F) + e_j \\ &= \lambda_j' F' + e_j \end{aligned}$$

$$\begin{aligned} d_1 &= F' + e_1 \\ d_2 &= \lambda_2' F' + e_2 \\ d_3 &= \lambda_3' F' + e_3 \\ &\vdots \end{aligned}$$

$\boldsymbol{\Sigma}$ under the surrogate model

Covariance structure equations

$$\boldsymbol{\Sigma} = \left(\begin{array}{rrr} \phi+\omega_1 & \lambda_2\phi & \lambda_3\phi \\ \lambda_2\phi & \lambda_2^2\phi+\omega_2 & \lambda_2\lambda_3\phi \\ \lambda_3\phi & \lambda_2\lambda_3\phi & \lambda_3^2\phi+\omega_3 \end{array} \right)$$

Solutions assuming λ_2 and λ_3 non-zero:

$$\begin{array}{rcl} \lambda_2 & = & \frac{\sigma_{23}}{\sigma_{13}} \\ \lambda_3 & = & \frac{\sigma_{23}}{\sigma_{12}} \\ \phi & = & \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \end{array}$$

We have another three-variable rule.

Combined three-variable identification rule

For the record

For a centered factor analysis model with a single factor, the parameters will be identifiable provided that

- There are at least three reference variables.
- Either the factor is standardized and the sign one factor loading is known, or else at least one factor loading equals one.
- Errors are independent of one another and of the factors.

Add more variables and more factors using other rules.

Two Versions of $\boldsymbol{\Sigma}$

For the surrogate model with a factor loading set to one

Under the Surrogate Model

$$\left(\begin{array}{ccc} \phi + \omega_1 & \lambda_2\phi & \lambda_3\phi \\ & \lambda_2^2\phi + \omega_2 & \lambda_2\lambda_3\phi \\ & & \lambda_3^2\phi + \omega_3 \end{array} \right)$$

Under the Original Model

$$\left(\begin{array}{ccc} \lambda_1^2\phi + \omega_1 & \lambda_1\lambda_2\phi & \lambda_1\lambda_3\phi \\ & \lambda_2^2\phi + \omega_2 & \lambda_2\lambda_3\phi \\ & & \lambda_3^2\phi + \omega_3 \end{array} \right)$$

	Value under model	
Function of $\boldsymbol{\Sigma}$	Surrogate	Original
$\frac{\sigma_{23}}{\sigma_{13}}$	λ_2	$\frac{\lambda_2}{\lambda_1}$
$\frac{\sigma_{23}}{\sigma_{12}}$	λ_3	$\frac{\lambda_3}{\lambda_1}$
$\frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}$	ϕ	$\lambda_1^2\phi$

Under the surrogate model

- It looks like λ_j is identifiable, but actually it's λ_j/λ_1.
- Estimates of λ_j for $j \neq 1$ are actually estimates of λ_j/λ_1.
- It looks like ϕ is identifiable, but actually it's $\lambda_1^2\phi$.
- ϕ is being expressed as a multiple of λ_1^2.
- Estimates of ϕ are actually estimates of $\lambda_1^2\phi$.

- Make d_1 the clearest representative of the factor.

Add an observable variable to the surrogate model

- Parameters are all identifiable, even if the factor loading of the new variable equals zero.
- Equality restrictions on $\boldsymbol{\Sigma}$ are created, because we are adding more equations than unknowns.
- These equality restrictions apply to the original model.
- It is straightforward to see what the restrictions are, though the calculations can be time consuming.

Finding the equality restrictions

- Calculate $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.
- Solve the covariance structure equations explicitly, obtaining $\boldsymbol{\theta}$ as a function of $\boldsymbol{\Sigma}$.
- Substitute the solutions back into $\boldsymbol{\Sigma}(\boldsymbol{\theta})$.
- Simplify.

Example: Add a 4th variable

$$
\begin{aligned}
d_1 &= F + e_1 \\
d_2 &= \lambda_2 F + e_2 \\
d_3 &= \lambda_3 F + e_3 \\
d_4 &= \lambda_4 F + e_4
\end{aligned}
$$

$e_1, \ldots, e_4, F$ all independent
$Var(e_j) = \omega_j \quad Var(F) = \phi$
$\lambda_1, \lambda_2, \lambda_3 \neq 0$

There are 8 parameters.

Covariance Matrix

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \left(\begin{array}{cccc} \phi + \omega_1 & \lambda_2\phi & \lambda_3\phi & \lambda_4\phi \\ \lambda_2\phi & \lambda_2^2\phi + \omega_2 & \lambda_2\lambda_3\phi & \lambda_2\lambda_4\phi \\ \lambda_3\phi & \lambda_2\lambda_3\phi & \lambda_3^2\phi + \omega_3 & \lambda_3\lambda_4\phi \\ \lambda_4\phi & \lambda_2\lambda_4\phi & \lambda_3\lambda_4\phi & \lambda_4^2\phi + \omega_4 \end{array} \right)$$

Solutions

$$\lambda_2 = \frac{\sigma_{23}}{\sigma_{13}}$$
$$\lambda_3 = \frac{\sigma_{23}}{\sigma_{12}}$$
$$\lambda_4 = \frac{\sigma_{24}}{\sigma_{12}}$$
$$\phi = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}$$

Substitute

$$\begin{array}{rcl} \sigma_{12} & = & \lambda_2\phi \\ & = & \dfrac{\sigma_{23}}{\sigma_{13}} \dfrac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \\ & = & \sigma_{12} \end{array}$$

Substitute solutions into expressions for the covariances

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \left(\begin{array}{rrrr} \phi + \omega_1 & \lambda_2\phi & \lambda_3\phi & \lambda_4\phi \\ \lambda_2\phi & \lambda_2^2\phi + \omega_2 & \lambda_2\lambda_3\phi & \lambda_2\lambda_4\phi \\ \lambda_3\phi & \lambda_2\lambda_3\phi & \lambda_3^2\phi + \omega_3 & \lambda_3\lambda_4\phi \\ \lambda_4\phi & \lambda_2\lambda_4\phi & \lambda_3\lambda_4\phi & \lambda_4^2\phi + \omega_4 \end{array} \right)$$

Solutions

$\lambda_2 = \frac{\sigma_{23}}{\sigma_{13}}$

$\lambda_3 = \frac{\sigma_{23}}{\sigma_{12}}$

$\lambda_4 = \frac{\sigma_{24}}{\sigma_{12}}$

$\phi = \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}}$

$$\begin{array}{rcl} \sigma_{12} & = & \sigma_{12} \\ \sigma_{13} & = & \sigma_{13} \\ \sigma_{14} & = & \frac{\sigma_{24}\sigma_{13}}{\sigma_{23}} \\ \sigma_{23} & = & \sigma_{23} \\ \sigma_{24} & = & \sigma_{24} \\ \sigma_{34} & = & \frac{\sigma_{24}\sigma_{13}}{\sigma_{12}} \end{array}$$

Equality Constraints

There should be 4(4+1)/2 - 8 = 2

$$
\begin{aligned}
\sigma_{14}\sigma_{23} &= \sigma_{24}\sigma_{13} \\
\sigma_{12}\sigma_{34} &= \sigma_{24}\sigma_{13}
\end{aligned}
$$

These hold regardless of whether factor loadings are zero (1234).

$$\sigma_{12}\sigma_{34} = \sigma_{13}\sigma_{24} = \sigma_{14}\sigma_{23}$$

Add another 3-variable factor

- Identifiability is maintained.
- $\sigma_{14} = Cov(d_1, d_4) = Cov(F_1 + e_1, F_2 + e_4) = Cov(F_1, F_2) = \phi_{12}$
- Actually, $\sigma_{14} = \lambda_1 \lambda_4 \phi_{12}$ under the original model.
- The covariances of the surrogate model are just those of the surrogate model, multiplied by un-knowable positive constants.
- As more variables and more factors are added, all this remains true.

Comparing the surrogate models

- Either set variances of factors to one, or set one loading per factor to one.
- Both arise from a similar change of variables.
- $F_j' = \lambda_j F_j$ or $F_j' = \frac{1}{\sqrt{\phi_{jj}}} F_j$.
- *Meaning* of surrogate model parameters in terms of the original model is different except for the signs.
- Both surrogate models share the same equality constraints, and hence the same goodness of fit results for any given data set.
- These constraints are also true of the original model.

The Equivalence Rule

See text for proof

For a centered factor analysis model with at least one reference variable for each factor, suppose that surrogate models are obtained by either standardizing the factors, or by setting the factor loading of a reference variable equal to one for each factor. Then the parameters of one surrogate model are identifiable if and only if the parameters of the other surrogate model are identifiable.

Which re-parameterization is better?

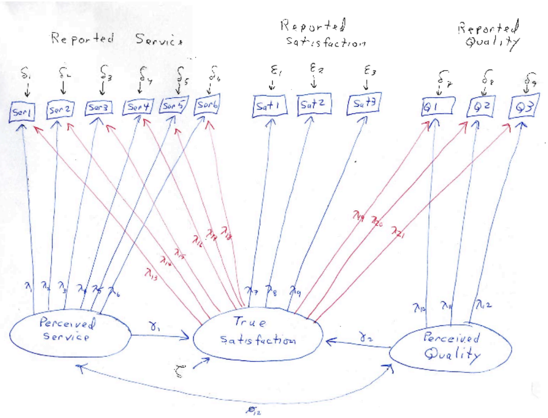
- Technically, they are equivalent.
- Interpretation of the surrogate model parameters is different.
- Standardizing the factors (Surrogate model 2A) is more convenient for estimating correlations between factors.
- But it's not robust to normality, so bootstrap it.
- Setting one loading per factor equal to one (Surrogate model 2B) is more convenient for estimating the relative sizes of factor loadings.
- Hand calculations and identifiability proofs with Surrogate model 2B can be easier.
- If there is a serious latent variable model, Surrogate model 2B is much easier to specify with lavaan.
- Mixing Surrogate model 2B with double measurement is natural.
- Don't do both restrictions for the same factor!

Why are we doing this? To buy identifiability.

- The parameters of the original model cannot be estimated directly. For example, maximum likelihood will fail because the maximum is not unique.
- The parameters of the surrogate models are identifiable (estimable) functions of the parameters of the true model.
- They have the same signs (positive, negative or zero) as the corresponding parameters of the true model.
- Hypothesis tests mean what you think they do.
- Parameter estimates can be useful if you know what the new parameters mean.

Crossover Patterns

- It is unfortunate when variables can only be infuenced by one factor. In fact, its unbelievable most of the time.
- A pattern like this would be nicer.



The Extra Variables Rule

A set of observable variables may be added to a measurement model whose parameters are **already identifiable**, provided

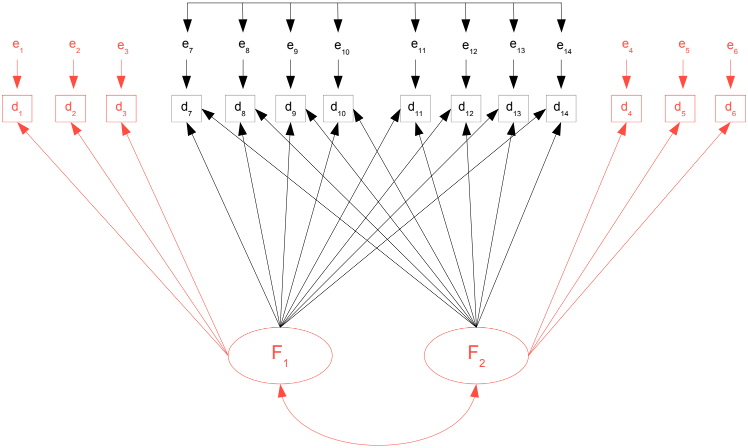
- There is a reference variable for each factor in the existing model.
- Error terms of the additional variables have zero covariance with the error terms of the reference variables in the existing model.
- Error terms of the additional variables have zero covariance with the factors.

Under these conditions,

- Straight arrows with factor loadings on them may point from each existing factor to each new variable.
- Error terms for the new set of variables may have non-zero covariances with each other, and with the error terms in the original model that do not belong to the reference variables.
- You don't need to include all such links.

All parameters of the new model are identifiable.

Adding Variables



Proof of The Extra Variables Rule

Extra variables are in $\mathbf{d}_3$

$$
\begin{aligned}
\mathbf{d}_1 &= \mathbf{F} + \mathbf{e}_1 \\
\mathbf{d}_2 &= \boldsymbol{\Lambda}_2\mathbf{F} + \mathbf{e}_2 \\
\mathbf{d}_3 &= \color{red}{\boldsymbol{\Lambda}_3}\mathbf{F} + \mathbf{e}_3
\end{aligned}
$$

$$
cov\left(\begin{array}{c} \mathbf{e}_1 \\ \hline \mathbf{e}_3 \\ \hline \mathbf{e}_3 \end{array}\right) = \left(\begin{array}{c|c|r} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} & \mathbf{0} \\ \hline & \boldsymbol{\Omega}_{22} & \color{blue}{\boldsymbol{\Omega}_{23}} \\ \hline & & \color{red}{\boldsymbol{\Omega}_{33}} \end{array}\right)
$$

Solving for $\mathbf{\Lambda}_3$, $\mathbf{\Omega}_{23}$ and $\mathbf{\Omega}_{33}$

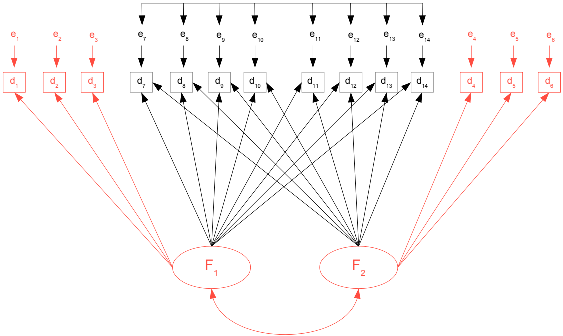
$$\mathbf{\Sigma} = cov \left(\begin{array}{c} \mathbf{d}_1 \\ \hline \mathbf{d}_2 \\ \hline \mathbf{d}_3 \end{array} \right) = \left(\begin{array}{c|c|c} \mathbf{\Sigma}_{11} & \mathbf{\Sigma}_{12} & \mathbf{\Sigma}_{13} \\ \hline & \mathbf{\Sigma}_{22} & \mathbf{\Sigma}_{23} \\ \hline & & \mathbf{\Sigma}_{33} \end{array} \right)$$

$$= \left(\begin{array}{c|r|r} \mathbf{\Phi} + \mathbf{\Omega}_{11} & \mathbf{\Phi}\mathbf{\Lambda}_2^\top & \mathbf{\Phi}\mathbf{\Lambda}_3^\top \\ \hline & \mathbf{\Lambda}_2\mathbf{\Phi}\mathbf{\Lambda}_2^\top + \mathbf{\Omega}_{22} & \mathbf{\Lambda}_2\mathbf{\Phi}\mathbf{\Lambda}_3^\top + \mathbf{\Omega}_{23} \\ \hline & & \mathbf{\Lambda}_3\mathbf{\Phi}\mathbf{\Lambda}_3^\top + \mathbf{\Omega}_{33} \end{array} \right)$$

Solve for a parameter and it becomes black.

$$\begin{aligned} \mathbf{\Lambda}_3 &= \mathbf{\Sigma}_{13}^\top \mathbf{\Phi}^{-1} \\ \mathbf{\Omega}_{33} &= \mathbf{\Sigma}_{33} - \mathbf{\Lambda}_3\mathbf{\Phi}\mathbf{\Lambda}_3^\top \\ \mathbf{\Omega}_{23} &= \mathbf{\Sigma}_{23} - \mathbf{\Lambda}_2\mathbf{\Phi}\mathbf{\Lambda}_3^\top \quad \blacksquare \end{aligned}$$

It's not so easy to put in all possible blue links



- Error terms of e_2, e_3, e_4 and e_5, can have covariances with $e_7 \ldots e_{14}$.
- Also, the Factor Model Combination Rule says e_2 may have covariances with e_4 and e_5, and e_3 may have covariances with e_4 and e_5.
- But there still cannot be covariances between e_2 and e_3, or between e_4 and e_5.
- It's hard to draw.

We can do a bit better

Adding $\mathbf{d}_3$ with the Extra Variables Rule

Instead of

$$\begin{aligned} \mathbf{d}_1 &= \mathbf{F} + \mathbf{e}_1 \\ \mathbf{d}_2 &= \boldsymbol{\Lambda}_2\mathbf{F} + \mathbf{e}_2 \\ \mathbf{d}_3 &= \boldsymbol{\Lambda}_3\mathbf{F} + \mathbf{e}_3 \end{aligned}$$

Could have

$$\begin{aligned} \mathbf{d}_1 &= \boldsymbol{\Lambda}_1\mathbf{F} + \mathbf{e}_1 \\ \mathbf{d}_2 &= \boldsymbol{\Lambda}_2\mathbf{F} + \mathbf{e}_2 \\ \mathbf{d}_3 &= \boldsymbol{\Lambda}_3\mathbf{F} + \mathbf{e}_3, \end{aligned}$$

with the $\boldsymbol{\Lambda}_1$ matrix $p \times p$ and invertable.

We have some identifiability rules for the measurement model

- Double Measurement rule.
- Scalar three-variable rule(s).
- The equivalence rule.
- Factor combination rule.
- Extra variable rule.
- Error-free rule.
- Need the 2-variable rules.
- Need the vector 3-variable rule.

Two-variable Rule

The parameters of a factor analysis model are identifiable provided

- There are two factors.
- There are two reference variables for each factor.
- For each factor, either the variance equals one and the sign of one factor loading is known, or the factor loading of at least one reference variable is equal to one.
- The two factors have non-zero covariance.
- Errors are independent of one another and of the factors.

Two-variable Rule

Two factors and two reference variables per factor

$$
\begin{aligned}
d_1 &= F_1 + e_1 \\
d_2 &= \lambda_2 F_1 + e_2 \\
d_3 &= F_2 + e_3 \\
d_4 &= \lambda_4 F_2 + e_4,
\end{aligned}
$$

with all expected values zero, $cov \left(\begin{array}{c} F_1 \\ F_2 \end{array} \right) = \mathbf{\Phi} = [\phi_{ij}]$, $Var(e_j) = \omega_j$, and the error terms independent of the factors and each other. An additional critical stipulation is that $Cov(F_1, F_2) = \phi_{12} \neq 0$.

Covariance matrix

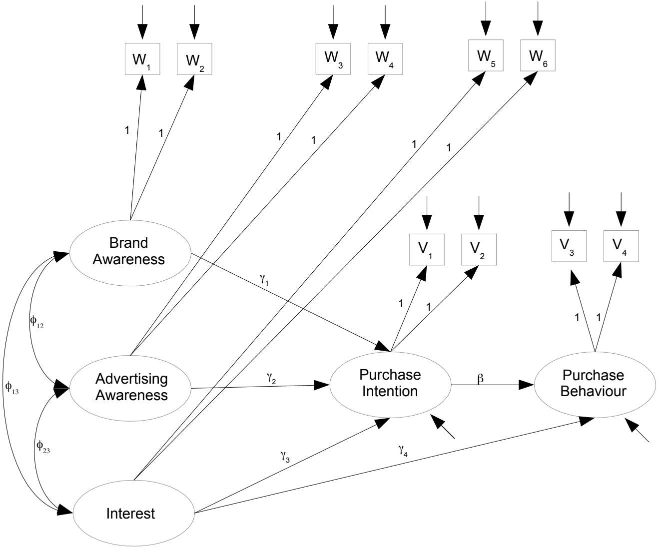
$$\begin{aligned}
\boldsymbol{\Sigma} &= cov\begin{pmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} & \sigma_{14} \\ & \sigma_{22} & \sigma_{23} & \sigma_{24} \\ & & \sigma_{33} & \sigma_{34} \\ & & & \sigma_{44} \end{pmatrix} \\
&= \begin{pmatrix} \phi_{11} + \omega_1 & \lambda_2\phi_{11} & \phi_{12} & \lambda_4\phi_{12} \\ & \lambda_2^2\phi_{11} + \omega_2 & \lambda_2\phi_{12} & \lambda_2\lambda_4\phi_{12} \\ & & \phi_{22} + \omega_3 & \lambda_4\phi_{22} \\ & & & \lambda_4^2\phi_{22} + \omega_4 \end{pmatrix}
\end{aligned}$$

Two-variable Addition Rule

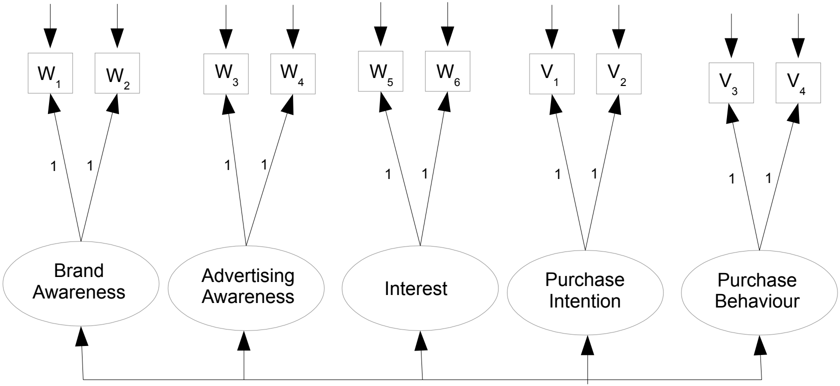
A factor with just two reference variables may be added to a measurement model whose parameters are identifiable, and the parameters of the combined model will be identifiable provided

- The errors for the two additional reference variables are independent of one another and of the error terms already in the model.
- Either the variance of the additional factor equals one and the sign of one factor loading is known, or the factor loading of at least one reference variable is equal to one.
- In the existing model with identifiable parameters,
  - There is at least one reference variable for each factor, and
  - At least one factor has a non-zero covariance with the new factor.

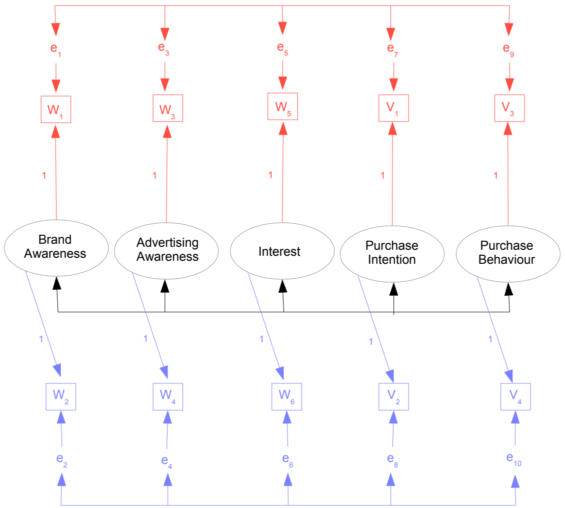
Brand Awareness Model 1



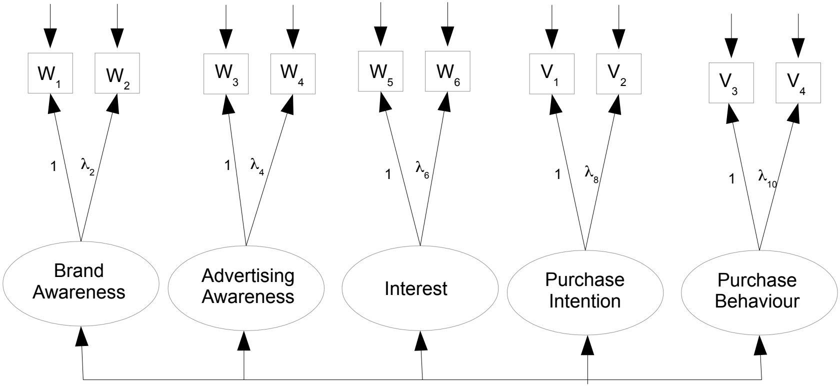
Brand Awareness Model 2



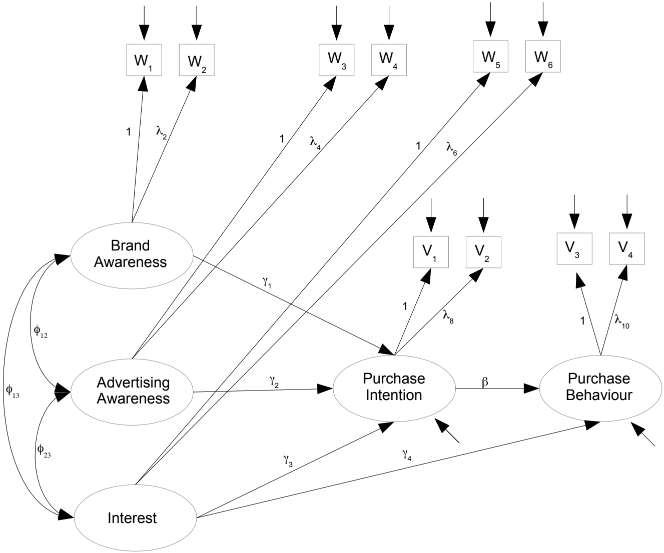
Brand Awareness Model 3



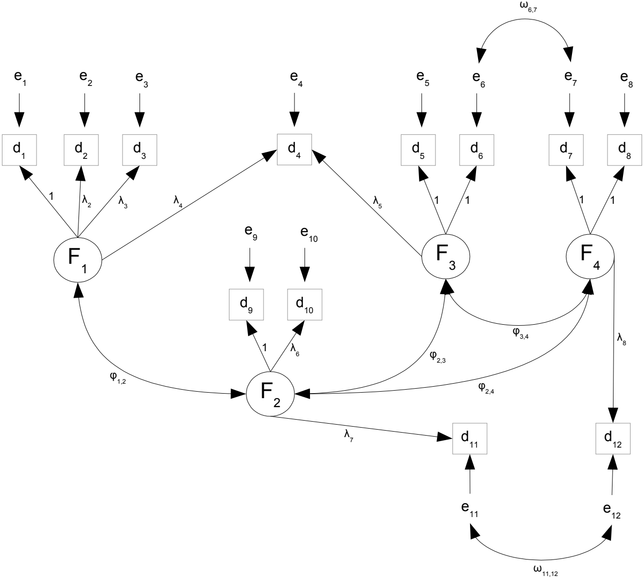
Brand Awareness Model 4



Brand Awareness Model 5



A big complicated measurement model



Vector 3-variable Rule

Let

$$\begin{aligned}\mathbf{d}_1 &= \mathbf{F} + \mathbf{e}_1 \\ \mathbf{d}_2 &= \boldsymbol{\Lambda}_2\mathbf{F} + \mathbf{e}_2 \\ \mathbf{d}_3 &= \boldsymbol{\Lambda}_3\mathbf{F} + \mathbf{e}_3\end{aligned}$$

where

- $\mathbf{F}$, $\mathbf{d}_1$ and $\mathbf{d}_2$ and $\mathbf{d}_3$ are all $p \times 1$.
- $\boldsymbol{\Lambda}_2$ and $\boldsymbol{\Lambda}_3$ have inverses.
- $cov(\mathbf{F}, \mathbf{e}_j) = cov(\mathbf{e}_i, \mathbf{e}_j) = \mathbf{O}$

Then all the parameters are identifiable.

Proof of Vector 3-variable Rule

Need to identify $\mathbf{\Phi}$, $\mathbf{\Lambda}_2$, $\mathbf{\Lambda}_3$, $\mathbf{\Omega}_{11}$, $\mathbf{\Omega}_{22}$, $\mathbf{\Omega}_{33}$

$$\begin{aligned} \mathbf{d}_1 &= \mathbf{F} + \mathbf{e}_1 \\ \mathbf{d}_2 &= \mathbf{\Lambda}_2\mathbf{F} + \mathbf{e}_2 \\ \mathbf{d}_3 &= \mathbf{\Lambda}_3\mathbf{F} + \mathbf{e}_3 \end{aligned}$$

$$\mathbf{\Sigma} = \left(\begin{array}{c|r|r} \mathbf{\Phi} + \mathbf{\Omega}_{11} & \mathbf{\Phi}\mathbf{\Lambda}_2^\top & \mathbf{\Phi}\mathbf{\Lambda}_3^\top \\ \hline & \mathbf{\Lambda}_2\mathbf{\Phi}\mathbf{\Lambda}_2^\top + \mathbf{\Omega}_{22} & \mathbf{\Lambda}_2\mathbf{\Phi}\mathbf{\Lambda}_3^\top \\ \hline & & \mathbf{\Lambda}_3\mathbf{\Phi}\mathbf{\Lambda}_3^\top + \mathbf{\Omega}_{33} \end{array}\right)$$

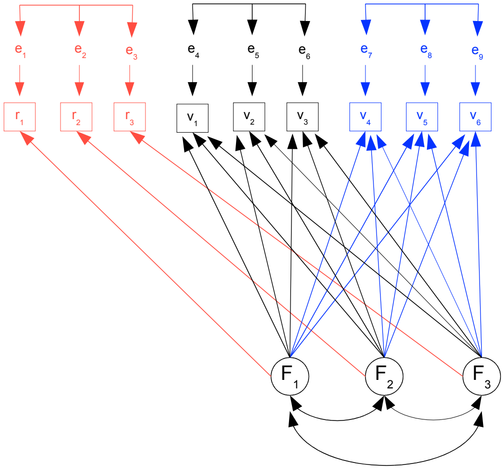
$$\begin{aligned} \mathbf{\Lambda}_2 &= \mathbf{\Sigma}_{23}\mathbf{\Sigma}_{13}^{-1} \\ \mathbf{\Phi} &= \mathbf{\Sigma}_{12}\left(\mathbf{\Lambda}_2^\top\right)^{-1} \\ \mathbf{\Lambda}_3 &= \left(\mathbf{\Phi}^{-1}\mathbf{\Sigma}_{13}\right)^\top \\ \mathbf{\Omega}_{11}, \mathbf{\Omega}_{22}, \mathbf{\Omega}_{33} & \quad \text{by subtraction.} \end{aligned}$$

■

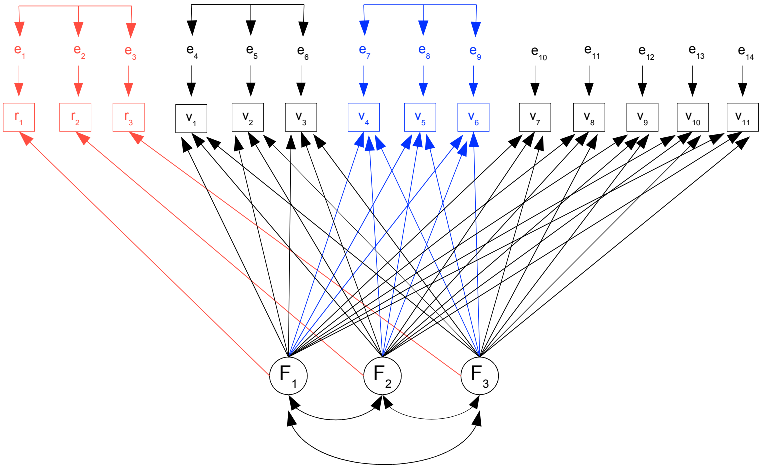
Don't need to include all possible links between error terms

By *omitting* certain permissible links between error terms, you can apply the Vector 3-variable Rule and the Additional Variables Rule to get some useful models that can be identified visually.

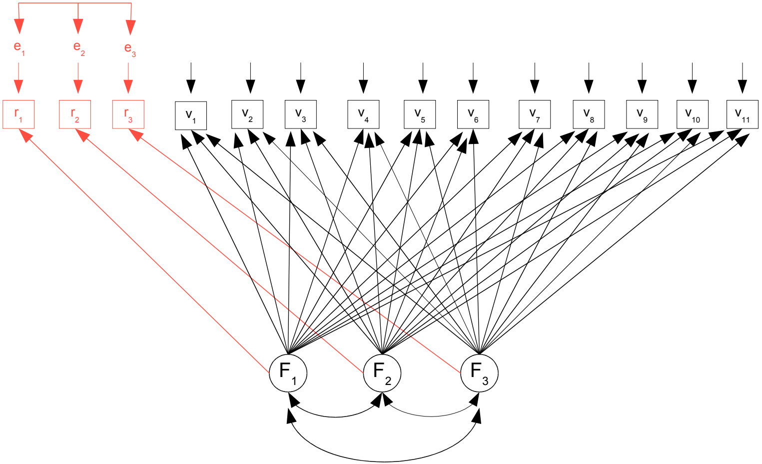
Vector 3-variable Rule



Vector 3-variable Rule and Additional Variables Rule



Drop some covariances between errors



Moral of the Story

With a reference variable for each factor, you can have a classical unrestricted exploratory factor analysis model for a set of additional observable variables.

- Factors may be correlated.
- Errors of reference variables may be correlated.

Student Mental Health Study Design

Factors are Anxiety and Depression

- Social variables
- Academic variables
- Psychological assessment variables

Social Variables

Using security camera recordings of students eating lunch in the cafeteria (with everyone's permission, of course), the investigators record four social behaviour variables during a designated twenty-minute period. Correlated errors within this set are very likely.

d_1: Speaking time (not on phone).

d_2: Listening time (head turned toward speaker).

d_5: Number of smiles/laughs while not on cell phone solo[2].

d_6: Time solo on cell phone.

[2] If two people are looking at a phone together, it's not "solo," and if they smile or laugh it would be counted

Academic variables

The following variables are obtained from school records. Measurement errors may not be correlated within this set, but we will be conservative, and assume they might be. In any case, it will be testable.

- d_3: Grade point average last academic session.
- d_4: Attendance last academic session.
- d_7: Hours per week playing school sports.
- d_8: Hours per week spent on extra-curricular activities, not including school sports.

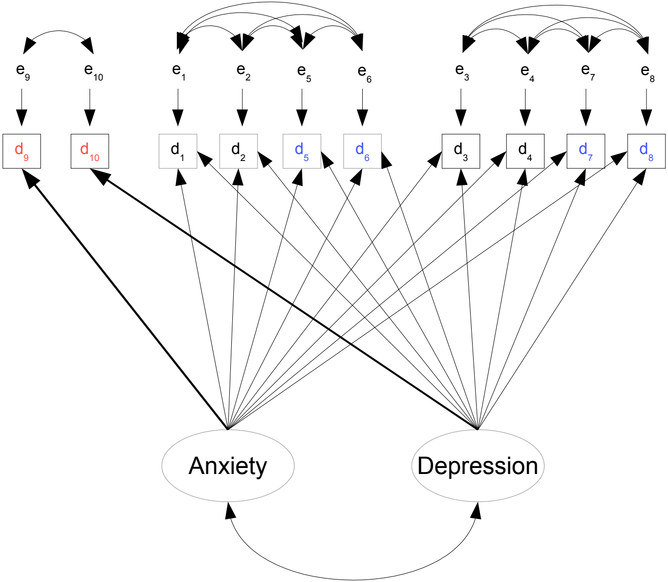
Psychological Assessment Variables

These are the reference variables

d_9: Clinical rating of anxiety.

d_{10}: Clinical rating of depression.

Model for the student mental health example



Moral of the Story

Divide variables into sets.

- Errors may be correlated within sets, but not between sets, as in double measurement.
- But sets need not measure the same things.
- One set consists of reference variables.
- In each set, at least as many variables as factors.
- Arrows from all factors to all variables that are not reference variables.

Bifactor Model from Uli Schimmack's blog

No reference variables!

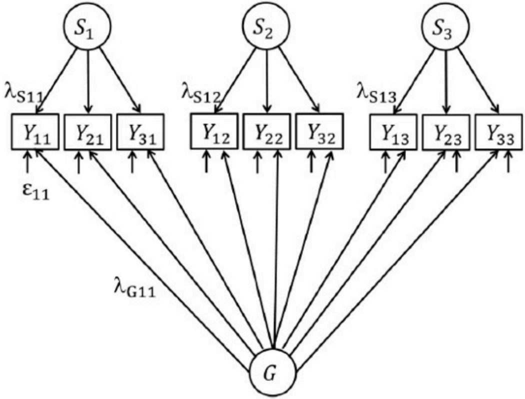


Figure 1. Traditional bifactor model with one general factor (G), three specific factors (S_k) and three observed variables Y_{ik} per domain. ε_{ik}: error variables, λ_{Gik}: G-factor loadings, λ_{Sik}: specific factor loadings $k = 1, \ldots, K$; K: number of domains; $i = 1, \ldots, I_k$; I_k: number of indicators i belonging to domain k. For simplicity, not all parameters and variables are labeled.

Bifactor Example

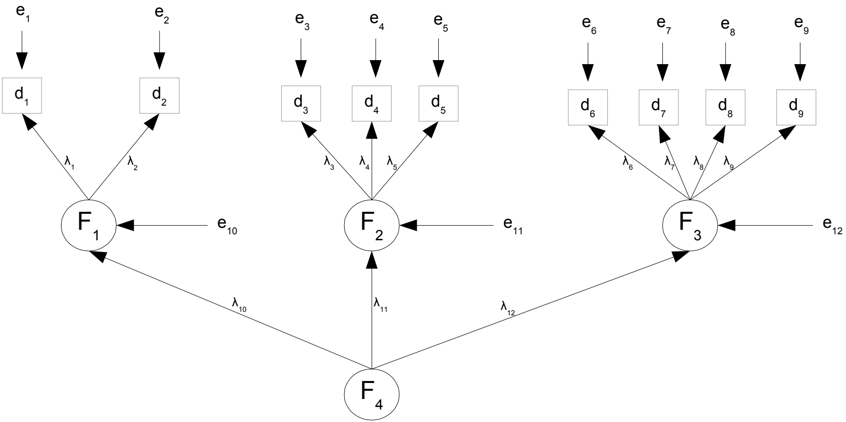
TABLE 1
Revised Child Anxiety and Depression Scale (*RCADS-15*)
Abbreviated Item Content and Proposed Disorder

1	*SAD*	Scared if I have to sleep on own
2	*SAD*	Afraid of on my own at home
3	*SAD*	Afraid of crowded places
4	*GAD*	Something will happen to family
5	*GAD*	Something bad will happen
6	*GAD*	I think about death
7	*PD*	Tremble or shake
8	*PD*	Suddenly become dizzy or faint
9	*PD*	Suddenly get a scared feeling
10	*SOC*	Worry when done poorly
11	*SOC*	Worry what other people think of me
12	*SOC*	Fool of myself in front of people
13	*OCD*	Have to think special thoughts
14	*OCD*	Do things over and over again
15	*OCD*	Do things in just the right way

Note. *SAD* = separation anxiety disorder; *GAD* = generalized anxiety disorder; *PD* = panic disorder; *SOC* = social phobia; *OCD* = obsessive-compulsive disorder.

From "The Rediscovery of Bifactor Measurement Models," by Steven P. Reise, *Multivariate Behavioral Research*, 47:667696, 2012 — Used without permission.

Second-order Factor Analysis



Copyright Information