Confirmatory Factor Analysis
Part Two

STA 2101: Fall/Winter 2019

See last slide for copyright information

THE TRUTH

(Well, closer to the truth, anyway)

Regression-like models

- Reality is massively non-linear.
- We can live with a linear approximation, as in multiple regression.
- All the model equations have unknown slopes and unknown intercepts.
- Like $D_1 = \lambda_{0,1} + \lambda_1 F_1 + e_1$
- Latent variables have unknown expected values and variances.
- Call this the "original model."

Identifiability

- If there are latent variables, the parameters of the original model are not identifiable.
- In a way, when we drop intercepts and ignore expected values, it’s like we are assuming all expected values = 0, “centering” the model.

- Original Model → Centered Model

Centering is a re-parameterization

- Not one-to-one.
- It reduces the dimension of the parameter space, helping with identifiability.
- Does not affect slopes, variances or covariances.
- *Meaning* is unaffected.

- What about $Var(F_j)=1$?

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 it's 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)=ϕ is an unknown parameter.

Centered Model

$$\begin{aligned} D_1 &= \lambda_1 F + e_1 \\ D_2 &= \lambda_2 F + e_2 \\ D_3 &= \lambda_3 F + e_3 \\ D_4 &= \lambda_4 F + e_4 \end{aligned}$$

$e_1, \ldots, e_4, F$ all independent

$Var(e_j) = \omega_j \quad Var(F) = \phi$

$\lambda_1, \lambda_2, \lambda_3 \neq 0$

Covariance Matrix

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

Passes the Counting Rule test with 10 equations in 9 unknowns

But for any c ≠ 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

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

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

You should be concerned!

- For any set of *true* parameter values, there are infinitely many *untrue* sets of parameter values that yield exactly the same Sigma and hence exactly the same probability distribution of the observable data.
- 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

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

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

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

So reliabilities are identifiable too.

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.

Testing Model Fit

- Note that all the equality constraints must involve only the covariances: σ_{ij} for $i \neq j$.
- In the original model, the covariances are all multiplied by the same non-zero constant.
- So, the equality constraints of the original model and the pretend model with $\phi=1$ are the same.
- The chi-square test for goodness of fit applies to the original model. This is a great relief!
- Likelihood ratio tests comparing full and reduced models are mostly valid without deep thought.
  - Equality of factor loadings is testable.
  - Could test H_0: $\lambda_4 = 0$, etc.

Re-parameterization

- The choice ϕ=1 is a very smart re-parameterization.
- It re-expresses the factor loadings as multiples of the square root of ϕ.
- 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

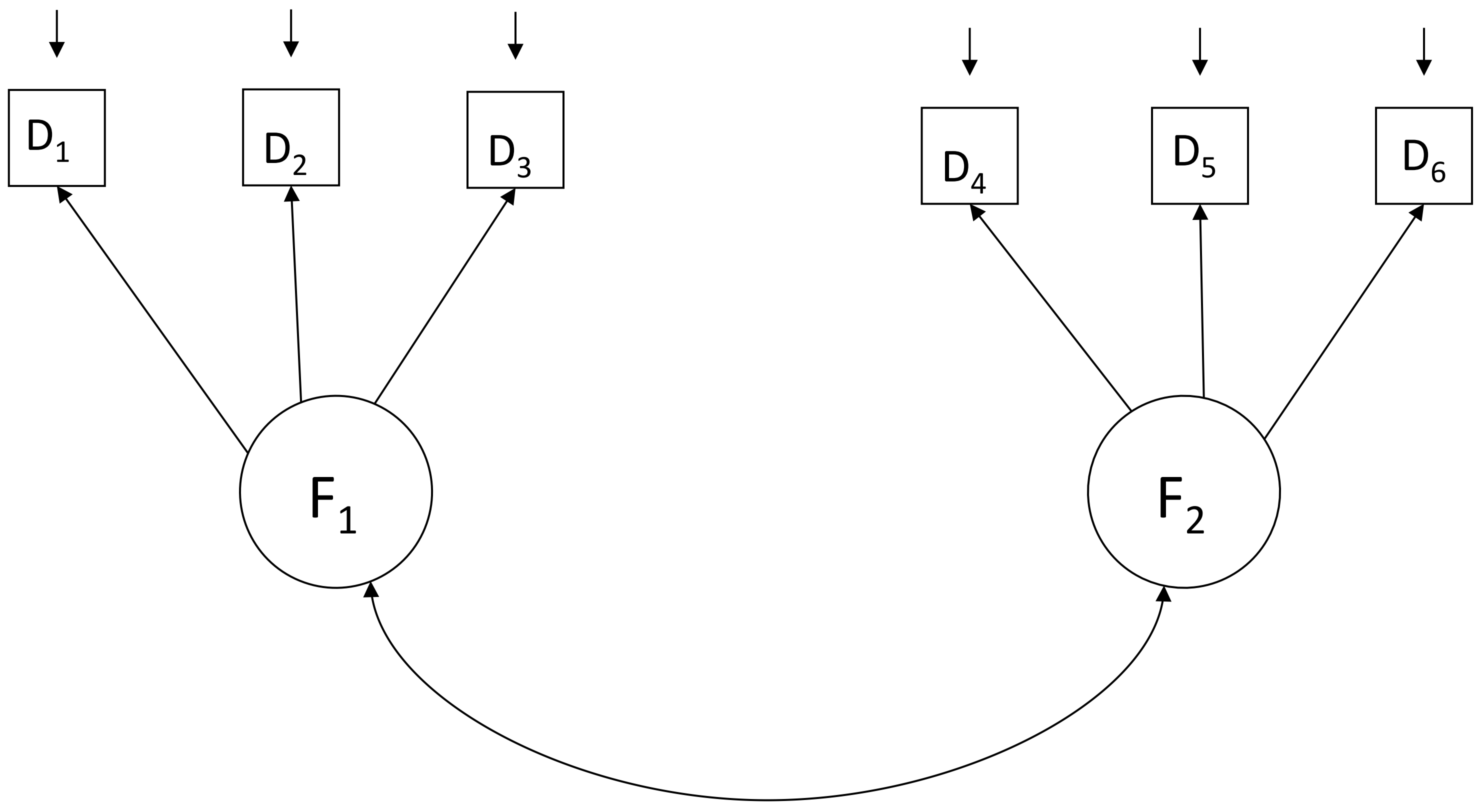
$\downarrow$

Surrogate model 1

$\downarrow$

Surrogate model 2

Add a factor to the centered model



Add a factor to the centered model

$$
\begin{aligned}
D_1 &= \lambda_1 F_1 + e_1 \\
D_2 &= \lambda_2 F_1 + e_2 \\
D_3 &= \lambda_3 F_1 + e_3 \\
D_4 &= \lambda_4 F_2 + e_4 \\
D_4 &= \lambda_5 F_2 + e_5 \\
D_6 &= \lambda_6 F_2 + e_6
\end{aligned}
$$

$$
cov \left(\begin{array}{c} F_1 \\ F_2 \end{array} \right) = \left(\begin{array}{cc} \phi_{11} & \phi_{12} \\ \phi_{12} & \phi_{22} \end{array} \right)
$$

$e_1, \ldots, e_6$ 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

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

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

$$\begin{array}{llll} \lambda_1' = c_1\lambda_1 & \lambda_2' = c_1\lambda_2 & \lambda_3' = c_1\lambda_3 & \phi_{11}' = \phi_{11}/c_1^2 \\ \lambda_4' = c_2\lambda_4 & \lambda_5' = c_2\lambda_5 & \lambda_6' = c_2\lambda_6 & \phi_{22}' = \phi_{22}/c_2^2 \\ \phi_{12}' = \frac{\phi_{12}}{c_1 c_2} & & & \end{array}$$

$$\omega_j' = \omega_j \text{ for } j = 1, \ldots, 6$$

Where $c_1 \neq 0$ and $c_2 \neq 0$

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 it's zero
- Sign is known if one factor loading from each set is known – say lambda1>0, lambda4>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 <u>correlation</u> between factors is identifiable!

The correlation between factors is identifiable

- Furthermore, it is the *same function of 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 equivalent to $Corr(F_1,F_2)$ under the original model.
- Estimates and tests of ϕ_{12} under the surrogate model apply to $\dfrac{\phi_{12}}{\sqrt{\phi_{11}}\sqrt{\phi_{22}}}$ under the original model.

Setting variances of factors to one

- Is a <u>very</u> smart re-parameterization.
- Is excellent when the interest is in correlations between factors.
- Allows estimation of classical path coefficients for the latent variable model.
- (That last remark was just for the record.)

Re-parameterization as a change of variables

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

- $Var(F_j') = 1$
- The new factor loading is in units of the standard deviation of F_j.
- This applies to all observable variables connected to F_j.
- Puts factor loadings for different factors on a common scale.

Covariances

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

- Covariances between factors in the surrogate model equal correlations in the original model.
- Latent variable parameters are strongly affected.
- Parameters in the latent surrogate model are the original parameters times positive constants.

What happens if there is a latent variable model?

$$\begin{aligned} Y_i &= \beta_1 X_i + \epsilon_i \\ Var(Y_i) &= \beta_1^2 \phi + \psi \end{aligned}$$

Standardize both X and Y.

$$\begin{array}{lcccccc} & \sqrt{\beta_1^2\phi+\psi}\left(\frac{1}{\sqrt{\beta_1^2\phi+\psi}}Y_i\right) & = & \beta_1\sqrt{\phi} & \left(\frac{1}{\sqrt{\phi}}X_i\right) & + & \epsilon_i \\ \Rightarrow & \left(\frac{1}{\sqrt{\beta_1^2\phi+\psi}}Y_i\right) & = & \left(\beta_1\sqrt{\frac{\phi}{\beta_1^2\phi+\psi}}\right) & \left(\frac{1}{\sqrt{\phi}}X_i\right) & + & \frac{\epsilon_i}{\sqrt{\beta_1^2\phi+\psi}} \\ & Y_i' & = & \beta_1' & X_i' & + & \epsilon_i' \end{array}$$

What does it mean?

$$
\begin{aligned}
Y_i' &= \beta_1' X_i' + \epsilon_i' \\
\beta_1' &= \beta_1 \sqrt{\frac{\phi}{\beta_1^2 \phi + \psi}}
\end{aligned}
$$

$$Cov(X_i', Y_i') = \beta_1' = Corr(X_i, Y_i)$$

Because covariances under the surrogate model equal correlations under the original model.

Factor Loadings are affected too

$$
\begin{aligned}
D_i &= \lambda Y_i + \cdots + e_i \\
&= \left(\lambda\sqrt{\beta_1^2\phi + \psi}\right)\left(\frac{1}{\sqrt{\beta_1^2\phi + \psi}}Y_i\right) + \cdots + e_i \\
&= \lambda' Y_i' + \cdots + e_i
\end{aligned}
$$

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.
- 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 one 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 λ_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

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

	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 $\lambda_1^2\phi$.

Everything is being expressed in terms of λ_1.

Make D_1 the clearest representative of the factor.

Add an observable variable

- Parameters are all identifiable, even if the factor loading of the new variable equals zero.
- Equality restrictions on 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

$Var(e_j) = \omega_j \quad Var(F) = \phi$

$\lambda_1, \lambda_2, \lambda_3 \neq 0$

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

Solutions

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

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

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

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

Substitute

$$\begin{aligned} \sigma_{12} &= \lambda_2\phi \\ &= \frac{\sigma_{23}}{\sigma_{13}} \frac{\sigma_{12}\sigma_{13}}{\sigma_{23}} \\ &= \sigma_{12} \end{aligned}$$

Substitute solutions into expressions for the covariances

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

Equality Constraints

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

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.
- The covariance $\phi_{12} = \sigma_{14}$
- 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.
- $F_j' = c_j F_j$, where $c_j > 0$.
- c_j is either a factor loading or one over a standard deviation.
- Interpretation of surrogate model parameters is different except for the sign.
- Mathematically the models are equivalent:

  Exchange λ_j and $\frac{1}{\sqrt{\phi_{jj}}}$.

- The true model and both surrogate models share the same equality constraints, and hence the same goodness of fit results for any given data set.

Which re-parameterization is better?

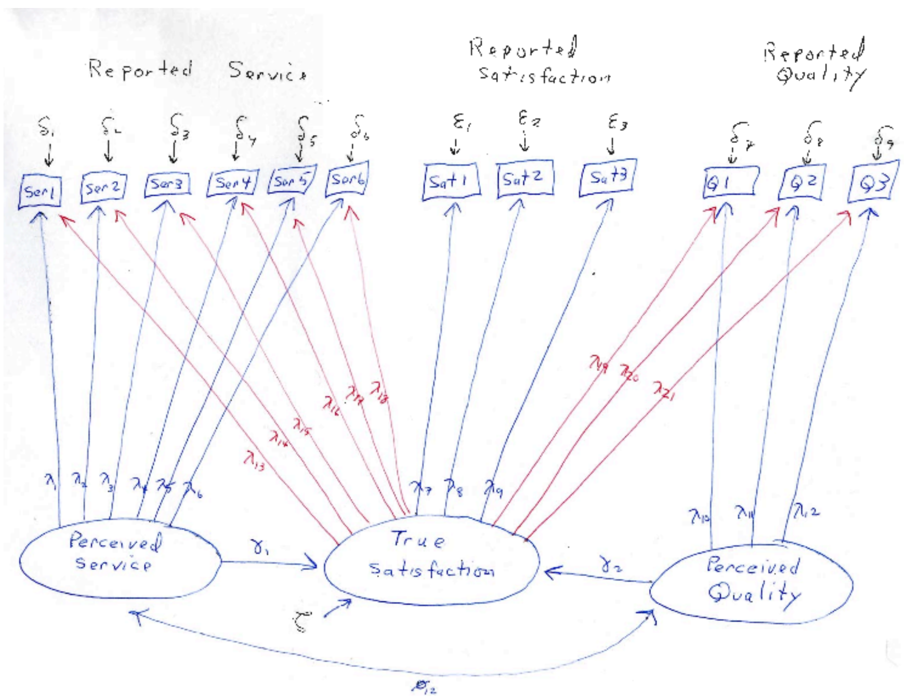
- Technically, they are equivalent.
- They both involve setting a single un-knowable parameter to one, for each factor.
- This seems arbitrary, but actually it results in a very good re-parameterization that preserves what is knowable about the true model.
- Standardizing the factors (Surrogate model 2A) is more convenient for estimating correlations between factors.
- Setting one loading per factor equal to one (Surrogate model 2B) is more convenient for estimating the relative sizes of factor loadings.
- Hand calculations with Surrogate model 2B can be easier.
- If there is a serious latent variable model, Surrogate model 2B is much easier to specify with SAS.
- Mixing Surrogate model 2B with double measurement is natural.
- Don’t do both restrictions for the same factor!

Why are we doing this?

- 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) of 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.

The Crossover Rule

- It is unfortunate that variables can only be caused by one factor. In fact, it’s unbelievable most of the time.
- A pattern like this would be nicer.



When you add a set of observable variables to a measurement model whose parameters are **already identifiable**

- Straight arrows with factor loadings on them may point from each existing factor to each new variable.
- You don't need to include all such arrows.
- Error terms for the new set of variables may have non-zero covariances with each other, but not with the error variances or factors of the original model.
- Some of the new error terms may have zero covariance with each other. It's up to you.
- All parameters of the new model are identifiable.

Proof

- Have a measurement (factor analysis) model with *p* factors and k_1 observable variables. The parameters are all identifiable.
- Assume that for each factor, there is at least one observable variable with a factor loading of one.
- If this is not the case, re-parameterize.
- Re-order the variables, putting the *p* variables with unit factor loadings first, in the order of the corresponding factors.

The first two equations belong to the initial model

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

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

$$
cov(\mathbf{F}) = \boldsymbol{\Phi}
$$

$$\boldsymbol{\Sigma} = \left(\begin{array}{c|c|c} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} & \boldsymbol{\Sigma}_{13} \\ \hline & \boldsymbol{\Sigma}_{22} & \boldsymbol{\Sigma}_{23} \\ \hline & & \boldsymbol{\Sigma}_{33} \end{array}\right)$$

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

Solve for it and it becomes black

$$\boldsymbol{\Lambda}_3 = \boldsymbol{\Sigma}_{13}^\top \boldsymbol{\Phi}^{-1}$$
$$\boldsymbol{\Omega}_{33} = \boldsymbol{\Sigma}_{33} - \boldsymbol{\Lambda}_3\boldsymbol{\Phi}\boldsymbol{\Lambda}_3^\top$$

Comments

- There are no restriction on the factor loadings of the variables that are being added to the model
- There are no restriction on the covariances of error terms for the new set of variables, except that they must not be correlated with error terms already in the model.
- This suggests a model building strategy. Start small, perhaps with 3 variables per factor. Then add the remaining variables – maximum flexibility.
- Could even fit the one-factor sub-models one at a time to make sure they are okay, then combine factors, then add variables.

Add an observed variable to the factors

- Often it's an observed exogenous variable (like sex or experimental condition) you want to be in a latent variable model.
- Suppose parameters of the existing (surrogate) factor analysis model (p factors) are all identifiable.
- X is independent of the error terms.
- Add a row (and column) to $\boldsymbol{\Sigma}$.
- Add p+1 parameters to the model.
- Say $Var(X)=\Phi_0$, $Cov(X,F_j)=\Phi_{0,j}$
- $D_k = \lambda_k F_j + e_k$, λ_k is already identified.
- $E(XD_k) = \lambda_k E(XF_j) + 0 = \lambda_k \Phi_{0,j}$
- Solve for the covariance.
- Do this for each factor in the model. Done.

We have some identification rules

- Double Measurement rule.
- Three-variable rule for standardized factors.
- Three-variable rule for unstandardized factors.
- Cross-over rule.
- Error-free rule.

Copyright Information