A single-factor Confirmatory Factor Analysis Model

(Starting simply)

Standardized

$$Z_1 = \lambda_1 F + e_1$$
$$Z_2 = \lambda_2 F + e_2$$
$$Z_3 = \lambda_3 F + e_3$$

$V(F) = V(Z_1) = V(Z_2) = V(Z_3) = 1$

F, e_1, e_2, e_3 ind.

leave it

$$\Sigma = \begin{array}{c|ccc} & Z_1 & Z_2 & Z_3 \\ \hline Z_1 & 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 \\ Z_2 & & 1 & \lambda_2\lambda_3 \\ Z_3 & & & 1 \end{array}$$

$V(Z_1) = 1 = \lambda_1^2 + V(e_1)$

$\Rightarrow V(e_1) = 1 - \lambda_1^2$

etc

$$\Theta = (\lambda_1\ \lambda_2\ \lambda_3)$$

- What does the parameter space Θ look like? It's a cube.

Are the parameters identifiable?

- If only one is zero, you can tell. For ex if $\lambda_1 = 0$ but not λ_2 or λ_3, then λ_1 is ident but $\sigma_{23} = \lambda_2\lambda_3$ & they are not.

λ_1 is identifiable on $\{\Theta \in \Theta : \lambda_1 = 0, \lambda_2 \neq 0, \lambda_3 \neq 0\}$

- Clearly we must examine identifiability at different points, regions of the

Suppose none of the factor loadings $= 0$

$$\frac{\sigma_{12}\,\sigma_{13}}{\sigma_{23}} = \frac{\lambda_1 \lambda_2 \lambda_1 \lambda_3}{\lambda_2 \lambda_3} = \lambda_1^2$$

$$\frac{\sigma_{12}\,\sigma_{23}}{\sigma_{13}} = \lambda_2^2 \quad , \quad \frac{\sigma_{13}\,\sigma_{23}}{\sigma_{12}} = \lambda_3^2$$

The squared factor loadings are identifiable, but not the factor loadings. Look at Σ. Replace λ_j by $-\lambda_j$ for $j = 1, 2, 3$ & get the same Σ.

Solution: Decide on the sign of one factor loading
(What if $Z_1 Z_2 Z_3$ are all math tests? Is F math ability or math inability? You decide)

For ex, if $\lambda_1 > 0$, you can tell the signs of λ_2 & λ_3 right away from Σ.

Additional points

- How many equality constraints? Zero
- How many inequality constraints; I say <u>one</u>
  $\sigma_{12}\,\sigma_{13}\,\sigma_{23} = 0$ These guys are slippery.
  No, there are more! If $\sigma_{23} \neq 0$, $\sigma_{12}\sigma_{13}/\sigma_{23} < 1$ etc.
- Expect 2 maxima of the likelihood function. Which one you find depends on where you start.

c

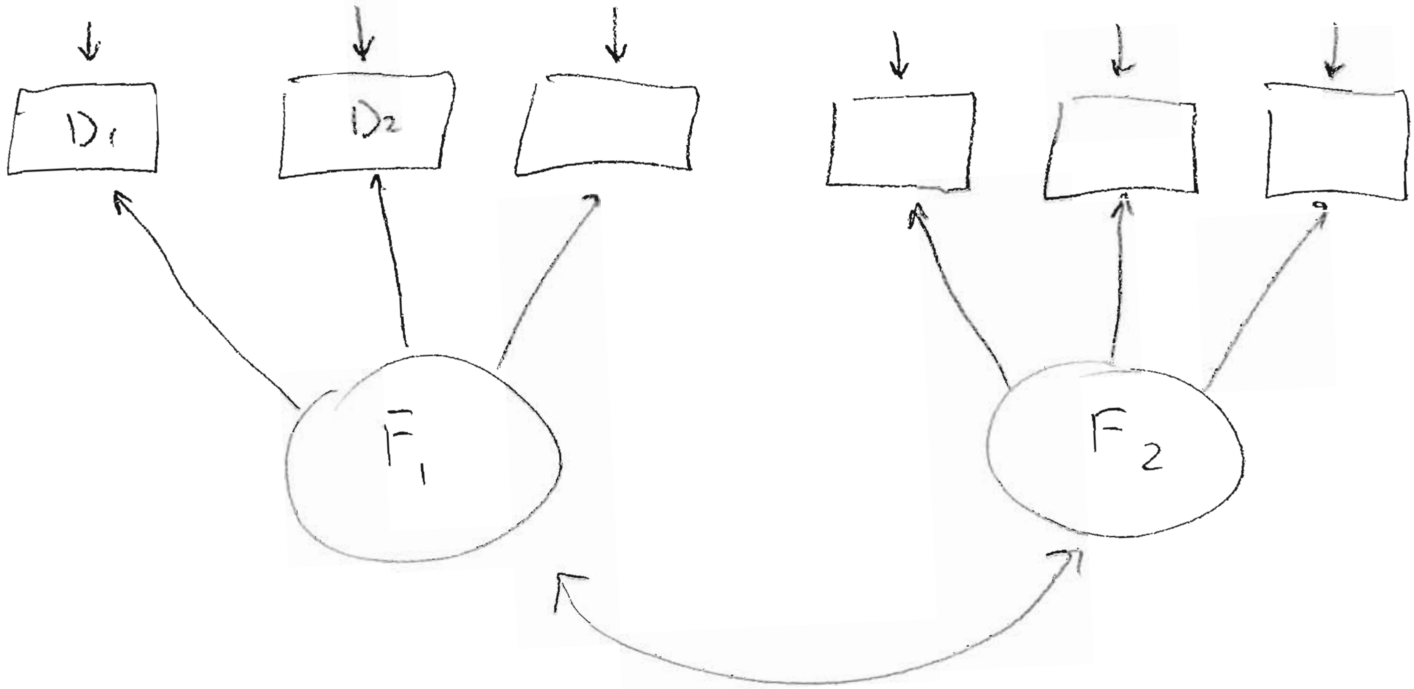
Add another variable : $X_4 = \lambda_4 F + e_4$

$$\Sigma = \begin{array}{|c|c|c|c|} \hline 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4 \\ \hline & 1 & \lambda_2\lambda_3 & \lambda_2\lambda_4 \\ \hline & & 1 & \lambda_3\lambda_4 \\ \hline & & & 1 \\ \hline \end{array}$$

- Parameters will all be identifiable as long as 3 out of 4 loadings are non-zero (and sign of one is known)
- For ex, if $\lambda_1 = 0$, top row $= 0$ and you can get $\lambda_2 \lambda_3 \lambda_4$ as before
- For 5 vars, two can be zero, etc.
- How many equality restrictions? $6 - 4 = 2$
- Inequality restrictions?: It's like an Easter egg hunt.

(d)

Now add another Factor



$$D_1 = \lambda_1 F_1 + e_1$$

$$\vdots$$

$$D_6 = \lambda_6 F_2 + e_6$$

$$\Sigma = \begin{array}{c|c|c|c|c|c|c}
 & D_1 & D_2 & D_3 & D_4 & D_5 & D_6 \\
\hline
D_1 & 1 & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4\phi_{12} & & \\
\hline
D_2 & & 1 & \lambda_2\lambda_3 & & \text{etc} & \\
\hline
D_3 & & & 1 & & & \\
\hline
D_4 & & & & 1 & \lambda_4\lambda_5 & \lambda_4\lambda_6 \\
\hline
D_5 & & & & & 1 & \lambda_5\lambda_6 \\
\hline
D_6 & & & & & & 1
\end{array}$$

Supposing you know one sign in each set,

- Identify λ_1 from set 1
- Ident λ_2 (set 2)
- Solve σ_{14} for ϕ_{12}

What if you added more vars?

" " " " " Factors?

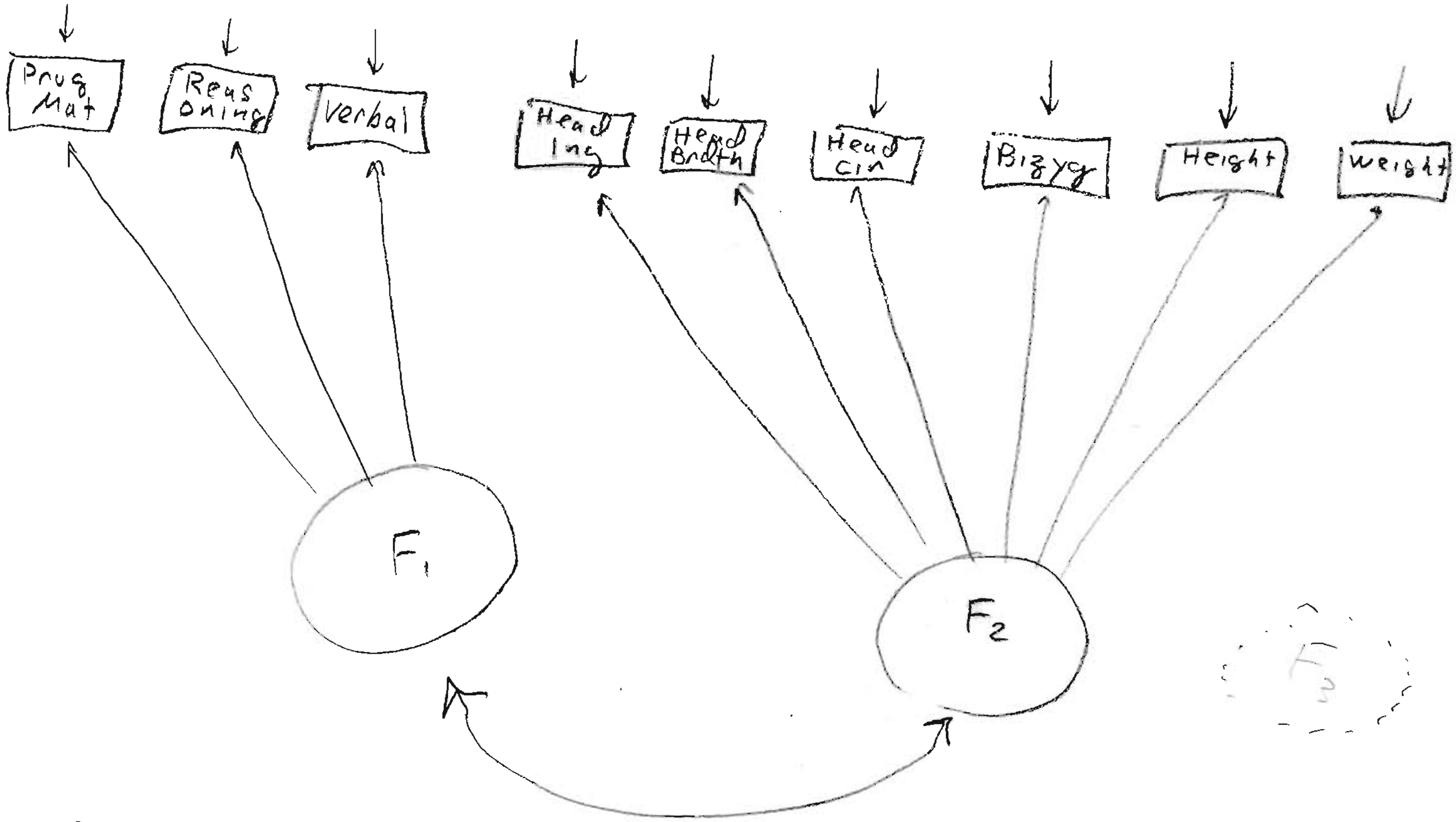
What if the obs vars were just centered, not standardized?

Rule

For a factor analysis model ~~with standardized observed variables (classical)~~ the parameters will be identifiable if

- Errors are independent of one another & of the factors
- V(Factors) = 1
- Each obs var is a function of only one factor
- There are at least 3 vars with non-zero loadings per factor
- Sign of one non-zero loading per factor is known.

Tried a 2-factor confirmatory FA on the twin data



But it did not fit. Separate factor for height and weight is attractive, but there are only 2 variables

Just 2 variables for the second factor
Both have non-zero factor loadings, and the sign of one is known.

$$D_1 = \lambda_1 F_1 + e_1$$
$$D_2 = \lambda_2 F_1 + e_2$$
$$D_3 = \lambda_3 F_1 + e_3$$
$$D_4 = \lambda_4 F_2 + e_4$$
$$D_5 = \lambda_5 F_2 + e_5$$

$V(F_1) = V(F_2) = 1$
$Cov(F_1, F_2) = \phi_{12} \neq 0$
$e_1 \cdots e_5$ ind & ind of F_1 & F_2
Say $\lambda_1 > 0$, $\lambda_4 > 0$

$$\Sigma = \begin{array}{c} \\ \end{array}\begin{array}{ccccc} 1 & 2 & 3 & 4 & 5 \\ \hline \checkmark & \lambda_1\lambda_2 & \lambda_1\lambda_3 & \lambda_1\lambda_4\phi_{12} & \lambda_1\lambda_5\phi_{12} \\ & \checkmark & \lambda_2\lambda_3 & \lambda_2\lambda_4\phi_{12} & \lambda_2\lambda_5\phi_{12} \\ & & - & \lambda_3\lambda_4\phi_{12} & \lambda_3\lambda_5\phi_{12} \\ & & & - & \lambda_4\lambda_5 \\ & & & & - \end{array}$$

$\lambda_1\ \lambda_2\ \lambda_3$ ident.

$$\frac{\sigma_{24}\,\sigma_{25}}{\sigma_{45}} = \frac{\lambda_2\lambda_4\phi_{12}\,\lambda_2\lambda_5\phi_{12}}{\lambda_4\lambda_5} = \lambda_2^2\phi_{12}^2$$

Since λ_2 is identifiable, this means ϕ_{12}^2 is ident.
And $\lambda_1 > 0$, $\lambda_4 > 0$, so sign of ϕ_{12} = Sign of σ_{14}.
And ϕ_{12} is ident.

Then $$\frac{\sigma_{45}}{\sigma_{34}} = \frac{\cancel{\lambda_4}\lambda_5}{\lambda_3\cancel{\lambda_4}\phi} = \frac{\lambda_5}{\lambda_3\phi_{12}}$$ Denom is ident, So λ_5 is ident. So λ_4 is ident.

Solving for $\omega_1 \cdots \omega_5$ is easy now.

So we have a two-variable rule (For standardized Factors)

We can add a factor that has only 2 vars provided

- Both loadings are non-zero
- Sign of one loading is known
- Has a non-zero covariance with a factor ~~the~~ with 3 variables.

So in the confirmatory factor analysis of the twin data, it is possible to have a 3d factor for height and weight

Why should we say that factors have variance one? "Because it's arbitrary" but saying it does not make it so. It's a re-parameterization, a semi-arbitrary restriction (reduction of the parameter space) to buy identifiability.

Truth

$D_1 = \lambda_1 F + e_1$

$D_2 = \lambda_2 F + e_2$

$D_3 = \lambda_3 F + e_3$

$D_4 = \lambda_4 F + e_4$.

$F, e_1 \cdots e_4$ ind

$V(e_j) = \omega_j$

$V(F) = \phi$

$\Sigma =$

$\lambda_1^2\phi + \omega_1$	$\lambda_1\lambda_2\phi$		
			$\lambda_4^2\phi + \omega_2$

10 eq in 9 unknowns, $\Theta = (\phi, \lambda_1\ \lambda_2\ \lambda_3\ \lambda_4\ \omega_1 \cdots \omega_4)$

Passes the test of the counting rule

j

But

	ϕ	λ_1	λ_2	λ_3	λ_4	ω_1	ω_2	ω_3	ω_4
θ_1	ϕ	λ_1	λ_2	λ_3	λ_4	ω_1	ω_2	ω_3	ω_4
θ_2	1	$\lambda_1\phi^{\frac{1}{2}}$	$\lambda_2\phi^{\frac{1}{2}}$	$\lambda_3\phi^{\frac{1}{2}}$	$\lambda_4\phi^{\frac{1}{2}}$	ω_1	ω_2	ω_3	ω_4

Yield the same covariance matrix

That's two θs yielding the same Σ, but there are ∞ many. Set $c > 0$

ϕ	λ_1	λ_2	λ_3	λ_4	ω_1	ω_2	ω_3	ω_4
c^2	$\frac{\lambda_1}{c}$	$\frac{\lambda_2}{c}$	$\frac{\lambda_3}{c}$	$\frac{\lambda_4}{c}$	ω_1	ω_2	ω_3	ω_4

ALL Yield

$$\Sigma = \begin{bmatrix} \lambda_1^2+\omega_1 & \lambda_1\lambda_2 & \text{etc} & \\ & \lambda_2^2+\omega_2 & & \\ & & \lambda_3^2+\omega_3 & \\ & & & \lambda_4^2+\omega_4 \end{bmatrix}$$

The choice $\phi = 1$ just sets $c = 1$: Convenient

(k)

You should be concerned!

Because the value of c is arbitrary except for $c > 0$, it would seem that the factor loadings could be anything. But it's not as bad as that

- Because $c > 0$, you can tell the true signs $+ - 0$ of the factor loadings (if you know one of them) with the arbitrary choice $\phi = 1$
- Same will be true of the ϕ_{ij} covariances
- Hypothesis testing is still on, but "estimation" is just a technical device

Testing the model

- Note that all the equality constraints must involve the covariances σ_{ij}, $i \neq j$
- In the true model, they are all multiplied by the same nonzero constant.
- So the equality constraints of the true model and the pretend model with $\phi = 1$ are the SAME.
- χ^2 test for goodness of fit applies to the true model. A great relief!