`cfa1`: Confirmatory Factor Analysis with a Single Factor

Stas Kolenikov

Department of Statistics
University of Missouri-Columbia

NASUG, Boston, MA, July 24, 2006

Outline

Factor analysis

If one has p variables $y_1, \ldots, y_p$, are there $q < p$ factors explaining most of the variability in y's?

- Exploratory factor analysis: find (simple) covariance structure in the data; a standard multivariate technique — see `[MV] factor`
- Confirmatory factor analysis: upon having formulated a theoretical model, see if it fits the data; estimate the parameters and assess goodness of fit. Simplest of structural equation models (SEM)
- Principal components analysis is neither of the above, but closer to EFA in spirit

Factor analysis

If one has p variables $y_1, \ldots, y_p$, are there $q < p$ factors explaining most of the variability in y's?

- Exploratory factor analysis: find (simple) covariance structure in the data; a standard multivariate technique — see `[MV] factor`
- Confirmatory factor analysis: upon having formulated a theoretical model, see if it fits the data; estimate the parameters and assess goodness of fit. Simplest of structural equation models (SEM)
- Principal components analysis is neither of the above, but closer to EFA in spirit

Factor analysis

If one has p variables $y_1, \dots, y_p$, are there $q < p$ factors explaining most of the variability in y's?

- Exploratory factor analysis: find (simple) covariance structure in the data; a standard multivariate technique — see `[MV] factor`
- Confirmatory factor analysis: upon having formulated a theoretical model, see if it fits the data; estimate the parameters and assess goodness of fit. Simplest of structural equation models (SEM)
- Principal components analysis is neither of the above, but closer to EFA in spirit

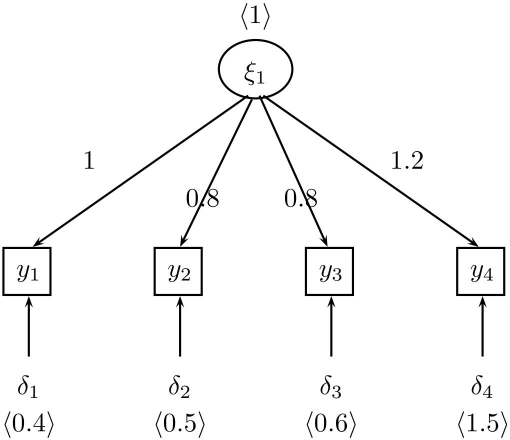
Factor analysis

If one has p variables $y_1, \ldots, y_p$, are there $q < p$ factors explaining most of the variability in y's?

- Exploratory factor analysis: find (simple) covariance structure in the data; a standard multivariate technique — see `[MV] factor`
- Confirmatory factor analysis: upon having formulated a theoretical model, see if it fits the data; estimate the parameters and assess goodness of fit. Simplest of structural equation models (SEM)
- Principal components analysis is neither of the above, but closer to EFA in spirit

Path diagrams

A typical CFA model looks like…



Equation notation

…which is the same as saying

$$
\begin{aligned}
&\xi_1 \text{ is unobserved,} && \mathbb{V}[\xi_1] = 1 = \phi \\
&y_1 = \lambda_1 \xi_1 + \delta_1, && \mathbb{V}[\delta_1] = 0.4 = \theta_1, \lambda_1 = 1, \\
&y_2 = \lambda_2 \xi_1 + \delta_2, && \mathbb{V}[\delta_2] = 0.5 = \theta_2, \lambda_2 = 0.8, \\
&y_3 = \lambda_3 \xi_1 + \delta_3, && \mathbb{V}[\delta_3] = 0.6 = \theta_3, \lambda_3 = 0.8, \\
&y_4 = \lambda_4 \xi_1 + \delta_3, && \mathbb{V}[\delta_4] = 1.5 = \theta_4, \lambda_4 = 1.2
\end{aligned}
$$

Likelihood formulation

$$y = \Lambda\xi_1 + \delta,$$

$$\mathbb{C}\mathrm{ov}[y] = \Lambda\phi\Lambda' + \Theta = \Sigma(\theta),$$

$$\ln L = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{n-1}{2}\operatorname{tr} S\Sigma^{-1}(\theta)$$

where S is the sample covariance matrix

Likelihood formulation

$$y = \Lambda\xi_1 + \boldsymbol{\delta},$$
$$\mathbb{C}\mathrm{ov}[y] = \Lambda\phi\Lambda' + \Theta = \Sigma(\boldsymbol{\theta}),$$
$$\ln L = -\frac{np}{2}\ln(2\pi) - \frac{n}{2}\ln|\Sigma| - \frac{n-1}{2}\operatorname{tr} S\Sigma^{-1}(\theta)$$

where S is the sample covariance matrix

Likelihood formulation

$$
\begin{aligned}
y &= \Lambda \xi_1 + \boldsymbol{\delta}, \\
\mathbb{C}\mathrm{ov}[y] &= \Lambda \phi \Lambda' + \Theta = \Sigma(\boldsymbol{\theta}), \\
\ln L &= -\frac{np}{2} \ln(2\pi) - \frac{n}{2} \ln |\Sigma| - \frac{n-1}{2} \operatorname{tr} S \Sigma^{-1}(\boldsymbol{\theta})
\end{aligned}
$$

where S is the sample covariance matrix

Identification conditions

Not all parameters are necessarily estimable...

- Number of parameters $\leq p(p+1)/2$
- $\mathbb{E}\,\xi_1$ is not identified, assumed zero
- Only $\lambda_k \phi^{1/2}$, or ratios λ_k/λ_j, are identified
  - Set $\phi = 1$
  - Set one of $\lambda_k = 1$

Identification conditions

Not all parameters are necessarily estimable…

- Number of parameters $\leq p(p+1)/2$
- $\mathbb{E}\,\xi_1$ is not identified, assumed zero
- Only $\lambda_k \phi^{1/2}$, or ratios λ_k/λ_j, are identified
    - Set $\phi = 1$
    - Set one of $\lambda_k = 1$

Identification conditions

Not all parameters are necessarily estimable...

- Number of parameters $\leq p(p+1)/2$
- $\mathbb{E}\,\xi_1$ is not identified, assumed zero
- Only $\lambda_k \phi^{1/2}$, or ratios λ_k/λ_j, are identified
  - Set $\phi = 1$
  - Set one of $\lambda_k = 1$

Stata implementation

- Stata's `ml model lf` structure
- Identification: by the first indicator, or by $\phi = 1$; implemented as `constraint`s supported by `ml`
- Improper solutions workarounds: what if $\hat{\theta}_k \leq 0$?
- Goodness of fit tests
- Corrections for multivariate kurtosis traditional for SEM literature (Satorra-Bentler standard errors and χ^2)

Stata implementation

- Stata's `ml model lf` structure
- Identification: by the first indicator, or by $\phi = 1$; implemented as `constraint`s supported by `ml`
- Improper solutions workarounds: what if $\hat{\theta}_k \le 0$?
- Goodness of fit tests
- Corrections for multivariate kurtosis traditional for SEM literature (Satorra-Bentler standard errors and χ^2)

Stata implementation

- Stata's `ml model lf` structure
- Identification: by the first indicator, or by $\phi = 1$; implemented as `constraint`s supported by `ml`
- Improper solutions workarounds: what if $\hat{\theta}_k \leq 0$?
- Goodness of fit tests
- Corrections for multivariate kurtosis traditional for SEM literature (Satorra-Bentler standard errors and χ^2)

Stata implementation

- Stata's `ml model lf` structure
- Identification: by the first indicator, or by $\phi = 1$; implemented as `constraint`s supported by `ml`
- Improper solutions workarounds: what if $\hat{\theta}_k \leq 0$?
- Goodness of fit tests
- Corrections for multivariate kurtosis traditional for SEM literature (Satorra-Bentler standard errors and χ^2)

Stata implementation

- Stata's `ml model lf` structure
- Identification: by the first indicator, or by $\phi = 1$; implemented as `constraint`s supported by `ml`
- Improper solutions workarounds: what if $\hat{\theta}_k \leq 0$?
- Goodness of fit tests
- Corrections for multivariate kurtosis traditional for SEM literature (Satorra-Bentler standard errors and χ^2)

Mata usage

- Normal likelihood, observation by observation; a lot of `st_view`'s and Cholesky decompositions. Earlier versions used `mkmat`... that was a disaster!
- Satorra-Bentler corrections:

$$\widehat{\text{acov}}(\hat\theta) = (n-1)^{-1}\left(\hat\Delta' V_n \hat\Delta\right)^{-1} \hat\Delta' V_n \Gamma_n V_n \hat\Delta \left(\hat\Delta' V_n \hat\Delta\right)^{-1}$$

$$\hat\Delta = \frac{\partial\sigma}{\partial\theta}\Big|_{\hat\theta}, V_n = 1/2\, D'(A_n^{-1} \otimes A_n^{-1})D,$$

$$A_n \xrightarrow{p} \Sigma, \operatorname{vec}\Sigma = D \operatorname{vech}\Sigma$$

$$\Gamma_n = \frac{1}{n-1}\sum_i (b_i - \bar b)(b_i - \bar b)'$$

$$b_i = (y_i - \bar y)(y_i - \bar y)'$$

- Mata functions for D (`vec`, `invvech`), $\hat\Delta$ (analytic derivatives, but with lots of matrix operations), b (`st_view`)

Mata usage

- Normal likelihood, observation by observation; a lot of `st_view`'s and Cholesky decompositions. Earlier versions used `mkmat`... that was a disaster!
- Satorra-Bentler corrections:

$$\widehat{\mathrm{acov}}(\hat{\boldsymbol{\theta}}) = (n-1)^{-1}\left(\hat{\Delta}' V_n \hat{\Delta}\right)^{-1} \hat{\Delta}' V_n \Gamma_n V_n \hat{\Delta} \left(\hat{\Delta}' V_n \hat{\Delta}\right)^{-1}$$

$$\hat{\Delta} = \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\theta}}\Big|_{\hat{\boldsymbol{\theta}}}, \; V_n = 1/2\, D'(A_n^{-1} \otimes A_n^{-1}) D,$$

$$A_n \xrightarrow{p} \Sigma, \mathrm{vec}\,\Sigma = D\,\mathrm{vech}\,\Sigma$$

$$\Gamma_n = \frac{1}{n-1} \sum_i (b_i - \bar{b})(b_i - \bar{b})'$$

$$b_i = (y_i - \bar{y})(y_i - \bar{y})'$$

- Mata functions for D (`vec`, `invvech`), $\hat{\Delta}$ (analytic derivatives, but with lots of matrix operations), b (`st_view`)

Mata usage

- Normal likelihood, observation by observation; a lot of `st_view`'s and Cholesky decompositions. Earlier versions used `mkmat`... that was a disaster!
- Satorra-Bentler corrections:

$$
\begin{aligned}
\widehat{\mathrm{acov}}(\widehat{\boldsymbol{\theta}}) &= (n-1)^{-1}\left(\hat{\Delta}' V_n \hat{\Delta}\right)^{-1} \hat{\Delta}' V_n \Gamma_n V_n \hat{\Delta} \left(\hat{\Delta}' V_n \hat{\Delta}\right)^{-1} \\
\hat{\Delta} &= \frac{\partial \boldsymbol{\sigma}}{\partial \boldsymbol{\theta}}\Big|_{\widehat{\boldsymbol{\theta}}}, \ V_n = 1/2\, D'(A_n^{-1} \otimes A_n^{-1}) D, \\
A_n &\xrightarrow{p} \Sigma, \mathrm{vec}\,\Sigma = D\,\mathrm{vech}\,\Sigma \\
\Gamma_n &= \frac{1}{n-1} \sum_i (b_i - \bar{b})(b_i - \bar{b})' \\
b_i &= (y_i - \bar{y})(y_i - \bar{y})'
\end{aligned}
$$

- Mata functions for D (`vec`, `invvech`), $\hat{\Delta}$ (analytic derivatives, but with lots of matrix operations), b (`st_view`)

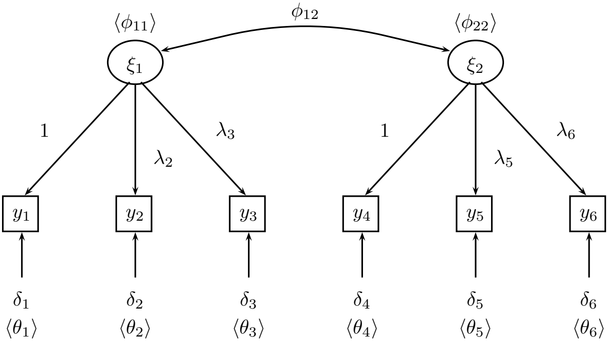
Syntax

`cfa1` *varlist* `[if ...]` `[in ...]`
`[[pweight=`*weight*`]],`
`unitvar free posvar`
`constraint(`*numlist*`)`
`cluster(`*varname*`)` `svy`
`robust` `vce(oim|opg|robust|sbentler)`
`from(`*starting values*`)` `level(#)`
ml options

Demonstration

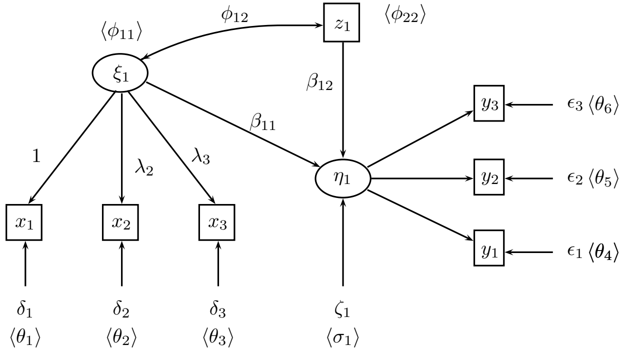
Do something!

Broader CFA models?




```
cfa (xi1:  y1 y2 y3) (xi2:  y4 y5 y6)
```

General structural equation models?




```
sem (xi1:  x1 x2 x3) (eta1 = .xi1 z1:  y1 y2 y3)
```

Parsing structural equation models is generally a nightmare...

GLLAMM

The minimal syntax for a CFA example for gllamm package

```
g long id = _n
reshape long y, i(id) j(k)
g byte _one = 1
tab k, gen(d)
eq main:  d1 d2 d3 d4
gllamm y d*, eq(main) i(id) nocons s(main)
```

Mplus

The minimal syntax for Mplus SEM package (text file input and output)

```
Title: Simple CFA example
Data:  File = cfa-example.txt;
       Type = individual;
Variable:
       Names = y1 y2 y3 y4;
Analysis:
       Type = general;
Model:
       xi by y1 y2 y3 y4;
```

SAS PROC CALIS

The minimal syntax for SAS PROC CALIS

```
proc calis data = cfa-example;
lineqs
   y1 =    f1 + e1,
   y2 = l2 f1 + e2,
   y3 = l3 f1 + e3,
   y4 = l4 f1 + e4;
std
   e1 = theta1,
   e2 = theta2,
   e3 = theta3,
   e4 = theta4,
   f1 = phi;
run;
```

Rigid variable names (F for factors/latent variables, E and D for errors and disturbances).

LISREL

```
TI PRELIS processing of CFA example
DA NI = 4 NO = 200 MI = -999
LA
   y1 y2 y3 y4
RA = cfa-example.txt
OU RA = cfa-example.psf

TI Estimation of CFA example
DA MA = CM NI = 4 NO = 200
RA = cfa-example.psf
LK xi
MO NK = 1 NX = 4
   LX = FU, FR
   TD = DI, FR
   FI LX(1,1) VA 1 LX(1,1)
PD
OU ME = ML EF
```

Download info

```
net from ///
http://www.missouri.edu/~kolenikovs/stata
net install cfa1

findit cfa1
```