Direct-Fitting Common Factor Analysis

Alwin Stegeman

a.w.stegeman@rug.nl
www.gmw.rug.nl/∼stegeman

VOC meeting 2015

university of groningen

faculty of behavioural and social sciences

Overview

Direct-Fitting Common Factor Analysis

1. The Common Factor Model
2. Factor Indeterminacy
3. Direct-Fitting Common Factor Analysis
4. Application
5. Conclusion

The Common Factor Model - formally

$$\boxed{X_j = \sum_{r=1}^{R} \lambda_{jr} F_r + u_j E_j \,, \qquad j = 1, \ldots, J}$$

X_j is observed variable j, F_r is common factor r, E_j is unique part j

λ_{jr} is loading of variable j on factor r, u_j is unique std j

$\phi_{rs} = \text{Corr}(F_r, F_s)$

Assumptions: X_j, F_r and E_j are standardized

$$\text{Corr}(E_j, F_r) = 0, \qquad \text{all } j, r$$

$$\text{Corr}(E_j, E_k) = 0, \qquad j \neq k$$

$$\text{Corr}(X_j, F_r) = \lambda_{jr} + \sum_{s \neq r} \lambda_{js} \phi_{rs}$$

The Common Factor Model - models

$$X_j = \sum_{r=1}^{R} \lambda_{jr} F_r + u_j E_j \,, \qquad\qquad j = 1, \ldots, J$$

Random Factor Model: F_r, E_j are random variables, λ_{jr}, u_j, ϕ_{rs} are fixed

Fixed Factor Model: E_j are random variables, F_r, λ_{jr}, u_j, ϕ_{rs} are fixed

Data Model: all of F_r, E_j, λ_{jr}, u_j, and ϕ_{rs} are fixed

Only fixed parameters are estimated

The Common Factor Model - estimation

$$X_j = \sum_{r=1}^{R} \lambda_{jr}\, F_r + u_j\, E_j\,, \qquad j = 1, \ldots, J$$

Random Factor Model: fit $\mathrm{Corr}(\mathbf{X}) \approx \mathbf{\Lambda}\,\mathbf{\Phi}\,\mathbf{\Lambda}^T + \mathbf{U}^2$, use:
MINRES (Harman & Jones, 1966), MLFA (Jöreskog, 1967), MRFA (Ten Berge & Kiers, 1991), ...

Fixed Factor Model: fit to observed data $\mathbf{X}$, use:
ML-ratio (McDonald, 1979), or first estimate random factor model and then 'estimate' factor scores as $\widehat{\mathbf{F}} = \mathbf{X}\,\mathbf{B}$ for some weights $\mathbf{B}$ (Grice, 2001)

Data Model: fit to observed data $\mathbf{X}$ (a.k.a. direct-fitting), use:
alternating LS (Kiers in Sočan, 2003; De Leeuw, 2004; Unkel & Trendafilov, 2010), DF-MRFA (Stegeman & Kiers, 2014), ...

Factor Indeterminacy

Data Model formulation: $\mathbf{X} = \mathbf{F}\,\mathbf{\Lambda}^T + \mathbf{E}\,\mathbf{U}$,

with $\mathbf{X}$ ($N \times J$), $\mathbf{F}$ ($N \times R$), $\mathbf{\Lambda}$ ($J \times R$), $\mathbf{E}$ ($N \times J$), and $\mathbf{U}$ diagonal ($J \times J$).

Assumptions: $N^{-1}\mathbf{E}^T\mathbf{F} = \mathbf{O}$, $N^{-1}\mathbf{E}^T\mathbf{E} = \mathbf{I}_J$, $N^{-1}\mathbf{F}^T\mathbf{F} = \mathbf{\Phi}$.

Factor scores $\mathbf{F}$ and unique part $\mathbf{E}$ are **not uniquely determined**!

$$\begin{aligned} \mathbf{F}, \mathbf{E} &= \text{determinate part} + \text{indeterminate part} \\ &= \text{regression on } \mathbf{X} + \text{residual} \end{aligned}$$

(regressions are $\mathbf{X}\,\mathbf{S}^{-1}\mathbf{\Lambda}\,\mathbf{\Phi}$ and $\mathbf{X}\,\mathbf{S}^{-1}\mathbf{U}$, with $\mathbf{S} = \text{Corr}(\mathbf{X})$)

Indeterminate parts of $\mathbf{F}$ and $\mathbf{E}$ are linked such that the assumptions hold.

Wilson (1928), Guttman (1955)

Factor Indeterminacy - minimal correlation

For two factors $\mathbf{f}_r$ and $\tilde{\mathbf{f}}_r$ with the same determinate part but different indeterminate parts, we have a minimal correlation

$$\text{Corr}(\mathbf{f}_r, \tilde{\mathbf{f}}_r) \geq 2\, R_r^2 - 1\,,$$

with $R_r^2 = (\boldsymbol{\Phi}\, \boldsymbol{\Lambda}^T \mathbf{S}^{-1} \boldsymbol{\Lambda}\, \boldsymbol{\Phi})_{rr}$ of the regression of $\mathbf{f}_r$ on $\mathbf{X}$ (Guttman, 1955).

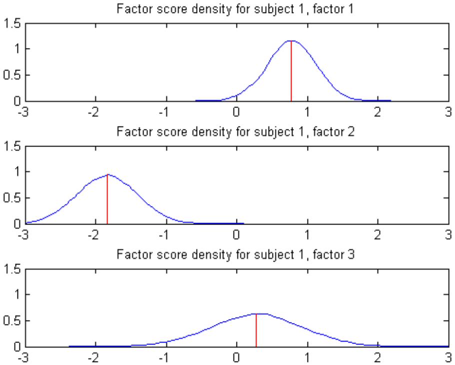
Factors are interpreted via $\text{Corr}(\mathbf{X}, \mathbf{F}) = \boldsymbol{\Lambda}\, \boldsymbol{\Phi}$, but what if the minimal correlation is very small or negative? (Guttman, 1955)

Minimal correlations should be reported when using Common Factor Analysis, or factor score std due to indeterminacy $\sqrt{1 - R_r^2}$.

Guttman (1955), Grice (2001)

Factor Indeterminacy - factor score probability densities

Factor score densities (subject 1) for noise-free data and $N = 100$. Based on simulating 10.000 realisations of indeterminate part, which contains a random matrix.



mean	std	min corr
0.78	0.34	0.77
-1.83	0.42	0.65
0.28	0.64	0.16

Mean = determinate part. Larger std for smaller minimal correlation.

Direct-Fitting Common Factor Analysis - models

Fitting the Data Model:

$$\text{DF-MINRES:} \quad \mathbf{X} = \overbrace{\mathbf{F}\,\boldsymbol{\Lambda}^T}^{\text{common part}} + \overbrace{\mathbf{E}\,\mathbf{U}}^{\text{unique part}} + \text{residual}$$

$$\text{DF-MRFA:} \quad \mathbf{X} = (\mathbf{F}\,\boldsymbol{\Lambda}^T + \text{residual}) + \mathbf{E}\,\mathbf{U}$$

Minimize ssq(residual) under constraints:

$N^{-1}\mathbf{E}^T\mathbf{F} = \mathbf{O}$, $N^{-1}\mathbf{E}^T\mathbf{E} = \mathbf{I}_J$, $N^{-1}\mathbf{F}^T\mathbf{F} = \mathbf{I}_R$

F and **E** have mean-zero columns

$N^{-1}\mathbf{E}^T(\text{common part}) = \mathbf{O}$

Direct-Fitting Common Factor Analysis - estimation

- DF-MINRES can be fitted by alternating LS, where each of $[\widehat{\mathbf{F}}\ \widehat{\mathbf{E}}]$, $\widehat{\mathbf{\Lambda}}$, and $\widehat{\mathbf{U}}$ is updated while keeping the others fixed (Kiers in Sočan, 2003; De Leeuw, 2004; Unkel & Trendafilov, 2010).
- DF-MRFA algorithm first computes $\widehat{\mathbf{U}}$ such that ssq(residual) is minimized (Ten Berge & Kiers, 1991). Next, $\widehat{\mathbf{E}}$ is computed, and $\widehat{\mathbf{F}}\,\widehat{\mathbf{\Lambda}}^T$ is a best rank-R approximation of $(\mathbf{X} - \widehat{\mathbf{E}}\,\widehat{\mathbf{U}})$ (Stegeman & Kiers, 2014).
- DF-MINRES and DF-MRFA: factor indeterminacy occurs in $\widehat{\mathbf{F}}$ and $\widehat{\mathbf{E}}$, but does not affect $\widehat{\mathbf{\Lambda}}$ and $\widehat{\mathbf{U}}$.
- DF-MRFA: proportion of explained common variance can be computed as $\text{ssq}(\widehat{\mathbf{F}}\,\widehat{\mathbf{\Lambda}}^T)/\text{ssq}(\widehat{\mathbf{F}}\,\widehat{\mathbf{\Lambda}}^T + \text{residual}) = \text{trace}(\widehat{\mathbf{\Lambda}}\widehat{\mathbf{\Lambda}}^T)/\text{trace}(\mathbf{S} - \widehat{\mathbf{U}}^2)$.

Direct-Fitting Common Factor Analysis - indeterminacy

- DF-MINRES and DF-MRFA: minimal correlation expressions of Guttman (1955) are valid, and also factor score due to indeterminacy.
- DF-MINRES and DF-MRFA: closed form expression for indeterminate part of $\widehat{\mathbf{F}}$ contains a random matrix, which makes it possible to sample it and obtain probablity densities for factor scores due to indeterminacy (under imperfect fit).

Stegeman & Kiers (2014)

Application - WISC-III

Wechsler Intelligence Scale for Children, 3rd edition (Wechsler, 1991).

$N = 280$, $J = 12$ subtests, $R = 4$ factors, Varimax rotation (Grice, 2001).

We use DF-MRFA and apply Varimax rotation to $\widehat{\boldsymbol{\Lambda}}$. Results are close to Grice (2001). Model fit is 95.7% and total ECV is 92.5%.

Factor score stds below are due to indeterminacy.

factor	ECV%	min corr	std
1	34.8	0.64	0.42
2	26.7	0.55	0.47
3	20.8	0.48	0.51
4	10.2	0.05	0.69

Application - WISC-III

subtest	$\mathbf{\Lambda}$				$\mathbf{U}^2$	ECV%
Picture Completion	0.26	**0.62**	0.16	0.03	0.48	91.7
Information	**0.75**	0.13	0.09	0.14	0.32	90.1
Coding	-0.02	0.15	**0.67**	0.08	0.48	92.0
Similarities	**0.71**	0.20	0.01	-0.00	0.42	93.9
Picture Arrangement	0.27	**0.43**	**0.44**	0.01	0.47	84.6
Arithmetic	0.32	0.04	0.30	**0.52**	0.48	88.2
Block Design	0.22	**0.72**	0.24	0.22	0.30	97.1
Vocabulary	**0.79**	0.21	0.10	0.07	0.32	99.2
Object Assembly	0.13	**0.76**	0.08	-0.09	0.35	93.9
Comprehension	**0.61**	0.14	0.14	0.24	0.50	93.7
Symbol Search	0.14	0.19	**0.77**	0.04	0.31	93.5
Digit Span	0.16	0.08	-0.04	**0.55**	0.62	87.8

Concluding Remarks

- Whichever factor model is used, some measure of factor indeterminacy should be reported for each factor.
- Factor indeterminacy is a controversial issue ("stop using the factor model" versus "assume infinitely many items measuring F, then the factor scores are unique in the limit"). (Steiger, 1979; Maraun, 1996; Mulaik, 2010).
- Factor score std values or probability densities due to indeterminacy are easier to grasp than minimal correlations (and less pessimistic).
- For $(\widehat{\mathbf{\Lambda}}, \widehat{\mathbf{U}})$ obtained from the random factor model, we can find all $(\widehat{\mathbf{F}}, \widehat{\mathbf{E}})$ with minimal ssq(residual) in DF-MINRES. Also, factor score std values due to factor indeterminacy can be simulated.
- DF-MRFA includes explained common variances and has estimation accuracy comparable to DF-MINRES and MINRES in simulations.

References

- De Leeuw, J. (2004). Least squares optimal scaling of partially observed linear systems. In: *Recent Developments on Structural Equation Models: Theory and Applications*, Eds. K. van Montfort, J. Oud & A. Satorra, pp. 121-134. Dordrecht: Kluwer Academic Publishers.
- Grice, J.W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*, 430–450.
- Guttman, L. (1955). The determinacy of factor score matrices with implications for five other basic problems of common factor theory. *The British Journal of Statistical Psychology*, *8*, 65–81.
- Harman, H.H., & Jones, W.H. (1966). Factor analysis by minimizing residuals (MINRES). *Psychometrika*, *31*, 351–368.
- Jöreskog, K.G. (1967). Some contributions to maximum likelihood factor analysis. *Psychometrika*, *32*, 443–482.
- Maraun, M.D. (1996). Metaphor taken as math: indeterminacy in the factor analysis model. With comments and reply. *Multivariate Behavioral Research*, *31*, 517–689.
- McDonald, R.P. (1979). The simultaneous estimation of factor loadings and scores. *British Journal of Mathematical and Statistical Psychology*, *32*, 212–228.
- Mulaik, S.A. (2010). *Foundations of Factor Analysis, 2nd edition*. Chapman & Hall, FL, USA.

References

- Sočan, G. (2003). *The Incremental Value of Minimum Rank Factor Analysis*. PhD Thesis, University of Groningen, The Netherlands.
- Steiger, J.H. (1979). Factor indeterminacy in the 1930s and the 1970s, some interesting parallels. *Psychometrika*, *44*, 157–167.
- Stegeman, A., & Kiers, H.A.L. (2014). Direct-fitting common factor analysis. Submitted.
- Ten Berge, J.M.F., & Kiers, H.A.L. (1991). A numerical approach to the approximate and the exact minimum rank of a covariance matrix. *Psychometrika*, *56*, 309–315.
- Unkel, S., & Trendafilov, N.T. (2010). Simultaneous parameter estimation in exploratory factor analysis: an expository review. *International Statistical Review*, *78*, 363-382.
- Wechsler, D. (1991). *WISC-III manual.* San Antonio, TX: Psychological Corporation.
- Wilson, E. (1928). On hierarchical correlation systems. *Proceedings of the National Academy of Sciences*, *14*, 283–291.