EFA in a CFA Framework

2012 San Diego Stata Conference

Phil Ender

UCLA Statistical Consulting Group
Institute for Digital Research & Education

July 26, 2012

Disclaimer

A researcher's attitudes and beliefs about factor analysis are largely determined by one's discipline or academic tribe.

Since I was raised in the Psychology Tribe I tend to have positive attitudes and beliefs concerning factor analysis.

A gross oversimplification of factor analysis

Factor analysis is concerned with the patterns of relationships between observed (manifest) variables and unobserved (latent) variables called factors.

Factor analysis comes in two major flavors:
1) Exploratory factor analysis (EFA), and
2) Confirmatory factory factor analysis (CFA).

EFA vs CFA

In exploratory factor analysis the researcher does not know the factor structure prior to running the analysis.

In confirmatory factor analysis the researcher "knows" the factor structure prior to the analysis and, in fact, sets which variables are indicators of which factor.

EFA in a CFA Framework

EFA in a CFA framework is a kind of a hybrid of EFA and CFA.

Uses CFA to obtain an EFA "like" solution.

EFA in a CFA framework imposes the same number of identifying restrictions on a CFA model as are found in an EFA model.

EFA in a CFA framework has the same fit as a maximum likelihood EFA solution.

Identifying restrictions

An EFA model with m factors will impose m^2 identifying restrictions.

Selecting identifying restrictions for EFA in a CFA framework:

1. Fix factor variances at 1
2. Select anchor items: variables with largest loading on each factor that have small loadings on the other factors.
3. Constrain the cross loadings for each anchor item to be zero for other factors.

Steps in the process

1. Obtain a rotated maximum likelihood factor analysis solution.
2. Identify an anchor item for each factor.
3. Set the cross factor loadings to zero for each anchor item.
4. Set the factor variances to one.
5. Run the **sem** command with the **standardized** option.

Step 1: Obtain rotated maximum likelihood solution

```
. use http://www.ats.ucla.edu/stat/data/efa_cfa, clear

. factor y1 y2 y3 y4 y5 y6, ml

. rotate, oblique quartimin normalize
```

Step 1: Rotated factor loadings results

```
LR test:
    2 factors vs. saturated: chi2(4) = 2.19 Prob>chi2 = 0.7015

  Rotated factor loadings (pattern matrix) and unique variances

    -------------------------------------------------
        Variable |  Factor1   Factor2 |   Uniqueness
    -------------+--------------------+--------------
              y1 |   0.6794    0.0117 |      0.5387
              y2 |   0.7657   -0.0228 |      0.4124
              y3 |   0.6972    0.0084 |      0.5141
              y4 |  -0.0073    0.6095 |      0.6282
              y5 |  -0.0281    0.7025 |      0.5048
              y6 |   0.0313    0.6086 |      0.6295
    -------------------------------------------------
```

Step 2: Identify an anchor item for each factor

```
LR test:
    2 factors vs. saturated: chi2(4) = 2.19 Prob>chi2 = 0.7015

  Rotated factor loadings (pattern matrix) and unique variances

    -------------------------------------------------
        Variable |  Factor1    Factor2 |   Uniqueness
    -------------+---------------------+-------------
              y1 |   0.6794     0.0117 |      0.5387
              y2 |  (0.7657)   -0.0228 |      0.4124
              y3 |   0.6972     0.0084 |      0.5141
              y4 |  -0.0073     0.6095 |      0.6282
              y5 |  -0.0281    (0.7025)|      0.5048
              y6 |   0.0313     0.6086 |      0.6295
    -------------------------------------------------
```

Variable **y2** will be the anchor for Factor 1 and **y5** will be the anchor for Factor 2.

Step 3: Set the cross factor loadings to zero for each anchor item

```
. sem (F1 -> y1 y2   y3 y4 y5@0 y6)    ///
      (F2 -> y1 y2@0 y3 y4 y5   y6)
```

Step 4: Set the factor variances to one

```
. sem (F1 -> y1 y2   y3 y4 y5@0 y6)    ///
      (F2 -> y1 y2@0 y3 y4 y5   y6) ,  ///
      variance(F1@1 F2@1)
```

Step 5: Run **sem** command with the **standardized** option

```
. sem (F1 -> y1 y2   y3 y4 y5@0 y6)    ///
      (F2 -> y1 y2@0 y3 y4 y5   y6) ,  ///
      variance(F1@1 F2@1) standardized
```

Step 5: Results

```
Iteration 0:   log likelihood = -5064.3487  (not concave)
Iteration 1:   log likelihood = -5005.6323  (not concave)
Iteration 2:   log likelihood = -4997.4943  (not concave)
Iteration 3:   log likelihood = -4982.0445  (not concave)
Iteration 4:   log likelihood = -4978.5317  (not concave)
[output omitted]
Iteration 141:  log likelihood =  2163798  (not concave)
Iteration 142:  log likelihood =  2163798  (not concave)
--Break--
```

Oops, **sem** would run forever without converging. We know the model is identified so we will try to to find some initial values.

Revised Step 5: Run **sem** with initial values

After a bit of experimenting using the **iterate** option the following initial values were selected.

```
. sem (F1 -> y1 y2   y3 y4 y5@0 ///
            (y6, init(0.0)))    ///
      (F2 -> y1 y2@0 y3 y4 y5  ///
            (y6, init(0.5))) ,  ///
      variance(F1@1 F2@1) standardized
```

Step 5: partial results

```
Endogenous variables

Measurement:  y1 y2 y3 y4 y5 y6

Exogenous variables

Latent:       F1 F2

Fitting target model:

Iteration 0:   log likelihood = -5467.1006  (not concave)
Iteration 1:   log likelihood = -5087.7064  (not concave)
[output omitted]
Iteration 8:   log likelihood = -4905.7634
Iteration 9:   log likelihood = -4905.7634

Structural equation model                   Number of obs =  500
Estimation method  = ml
Log likelihood     = -4905.7634
```

More Step 5: partial results edited for space

```
Standardized |      Coef.   Std. Err.
-------------+----------------------------------------
Measurement  |
y1        F1 |   .6814311   .0353949
          F2 |   .0319918   .0506257
       _cons |  -.0158254   .0447242
  -----------+----------------------------------------
y2        F1 |   .7665428   .0340075
       _cons |   .0183253   .0447251
     ----------+--------------------------------------
y3        F1 |   .6991976   .0352692
          F2 |   .0292474    .050601
       _cons |   .0248039   .0447282
  -----------+----------------------------------------
y4        F1 |   .0171268   .0517733
          F2 |   .6110693   .0447048
       _cons |  -.0156637   .0447241
  -----------+----------------------------------------
y5        F2 |   .7036822   .0455108
       _cons |  -.0121345    .044723
```

More Step 5: partial results edited for space con't

```
Standardized   |       Coef.   Std. Err.
---------------+----------------------------------------
y6          F1 |    .0557532   .0519712
            F2 |    .6112825     .04504
         _cons |    .0368424   .0447365
---------------+----------------------------------------
Variance e.y1 |    .5386683   .0475129
          e.y2 |    .4124122   .0521364
          e.y3 |    .5140572   .0485346
          e.y4 |    .6282406   .0540083
          e.y5 |    .5048314   .0640503
          e.y6 |    .6295414   .0541293
            F1 |           1  (constrained)
            F2 |           1  (constrained)
---------------+----------------------------------------
Covariance     |
   F1       F2 |   -.0926641   .0801889
--------------------------------------------------------
LR test of model vs. saturated: chi2(4) = 2.20,
                                Prob > chi2 = 0.6981
```

Summary of results

	EFA within CFA loadings		EFA rotated maximum likelihood loadings	
	F1	F2	Factor1	Factor2
y1	.6814311	.0319918	0.6794	0.0117
y2	.7665428	0	0.7657	-0.0228
y3	.6991976	.0292474	0.6972	0.0084
y4	.0171268	.6110693	-0.0073	0.6095
y5	0	.7036822	-0.0281	0.7025
y6	.0557532	.6112825	0.0313	0.6086

```
factor: LR test: 2 factors vs. saturated: chi2(4) = 2.19
                                          Prob>chi2 = 0.7015
   sem: LR test   of model vs. saturated: chi2(4) = 2.20
                                          Prob>chi2 = 0.6981
```

References

Adams, H. (1988) *The academic tribes.* Urbana and Chicago: University of Illinois Press.

Brown, T.A. (2006) *Confirmatory factor analysis for applied research.* New York: Guilford Press.

Jöreskog, K.G. (1969). A general approach to confirmatory maximum likelihood factor analysis. *Psychometrika*, *34*, 183-202.

Muthén, L.K. and Muthén, B.. (2009). *Mplus short courses topic 1: Exploratory factor analysis, confirmatory factor analysis, and structural equation modeling for continuous outcomes.* [PDF] URL http://www.statmodel.com/download/Topic%201.pdf