Confirmatory factor analysis of ordinal data without missing values

Contents

1. Introduction

Confirmatory factor analysis of ordinal data can be analyzed via one of two methods. The first option is to use full information maximum likelihood (FIML) using a probit, logit or other link function. The model can also be fitted to a matrix of polychoric correlations or a covariance matrix by using a fit function such as ULS or DWLS.

Using adaptive quadrature has certain disadvantages, as it does not provide an overall measure of fit nor does it five a completely standardized solution with standard errors. Using adaptive quadrature may also be very time-consuming if a large number of variables is used, especially if the number of factors is large.

Using polychoric correlations and their asymptotic covariance matric works well even for large numbers of factors and variables and also provides standardized solutions. In this example, we illustrate the use of polychoric correlations and DWLS, using the same data as was used to illustrate FIML.

The Eurobarometer Survey from 1192 asked respondents questions about science and technology. Here we look at 7 questions, with responses from UK citizens. The sample size is 392, and there are no missing values.

The response alternatives for the questions considered here were:

- strongly disagree (coded 1)
- disagree to some extent (coded 2)
- agree to some extent (coded 3)
- strongly agree (coded 4)

The questions we look at in this example are:

- COMFORT: science and technology are making our lives healthier, easier and more comfortable
- ENVIRON: scientific and technological research cannot play an important role in protecting the environment and repairing it
- WORK: the application of science and new technology will make work more interesting
- FUTURE: thanks to science and technology there will be more opportunities for the future generations
- TECHNOL: new technology does not depend on basic scientific research

- INDUSTRY: scientific and technological research do not play an important role in industrial development
- BENEFIT: the benefits of science are greater than any harmful effects it may have.

The first few lines of the data file **scitech.lsf** are shown below.

SCITECH.LSF

	COMFORT	ENVIRON	WORK	FUTURE	TECHNOL	INDUSTRY	BENEFIT
1	4.00	4.00	4.00	3.00	4.00	3.00	2.00
2	3.00	4.00	3.00	3.00	3.00	3.00	3.00
3	3.00	2.00	2.00	2.00	4.00	4.00	3.00
4	3.00	3.00	2.00	2.00	4.00	4.00	3.00
5	3.00	1.00	4.00	4.00	2.00	3.00	1.00
6	4.00	3.00	4.00	3.00	3.00	4.00	3.00
7	3.00	2.00	2.00	3.00	4.00	4.00	4.00
8	3.00	2.00	2.00	3.00	3.00	4.00	4.00
9	3.00	3.00	3.00	4.00	4.00	4.00	2.00
10	4.00	3.00	3.00	3.00	3.00	3.00	3.00

2. Estimation using polychoric correlations

The variables have already been declared as ordinal in nature in the data set, so no information on this is needed in the model specification. To request this analysis using SIMPLIS syntax, the file below may be used.

scitech5a.spl

```
Estimating the model using polychoric correlations
Raw Data from file scitech.lsf
Latent Variables Positive Negative
Relationships
COMFORT WORK FUTURE BENEFIT = Positive
COMFORT ENVIRON TECHNOL INDUSTRY    = Negative
Analyze Correlations
Robust Estimation
Method of Estimation: Diagonally Weighted Least Squares
Path Diagram
End of Problem
```

Alternatively, the analysis may be performed using the equivalent LISREL syntax.

scitech5b.lis

```
Estimating the model using polychoric correlations
da ni=7 ma=km
ra=SCITECH.lsf
mo nx=7 nk=2
lk
Positive Negative
fr lx(1,1) lx(1,2) lx(2,2) lx(3,1) lx(4,1) lx(5,2) lx(6,2) lx(7,1)
ro
pd
ou me=dwls
```

The first part of the output reports the univariate statistics. Note that there are three estimated thresholds for each of the variables. These are estimated from the marginal distributions.

Univariate Marginal Parameters

Variable	Mean	St. Dev.	Thresholds		
COMFORT	0.000	1.000	-2.234	-1.314	0.749
ENVIRON	0.000	1.000	-1.447	-0.514	0.450
WORK	0.000	1.000	-1.377	-0.428	1.079
FUTURE	0.000	1.000	-1.803	-0.774	0.691
TECHNOL	0.000	1.000	-1.686	-0.589	0.464
INDUSTRY	0.000	1.000	-1.951	-1.056	0.219
BENEFIT	0.000	1.000	-1.611	-0.500	0.845

Univariate Distributions for Ordinal Variables

COMFORT	Frequency	Percentage	Bar Chart
1	5	1.3	
2	32	8.2	
3	266	67.9	
4	89	22.7	

ENVIRON	Frequency	Percentage	Bar Chart
1	29	7.4	
2	90	23.0	
3	145	37.0	
4	128	32.7	

WORK	Frequency	Percentage	Bar Chart
1	33	8.4	
2	98	25.0	
3	206	52.6	
4	55	14.0	

FUTURE	Frequency	Percentage	Bar Chart
1	14	3.6	
2	72	18.4	
3	210	53.6	
4	96	24.5	

TECHNOL	Frequency	Percentage	Bar Chart
1	18	4.6	
2	91	23.2	
3	157	40.1	
4	126	32.1	

INDUSTRY	Frequency	Percentage	Bar Chart
1	10	2.6	
2	47	12.0	
3	173	44.1	
4	162	41.3	

BENEFIT	Frequency	Percentage	Bar Chart
1	21	5.4	
2	100	25.5	
3	193	49.2	
4	78	19.9	

The estimated polychoric correlation matrix is also given in the output file

```
      Correlation Matrix

             COMFORT    ENVIRON       WORK     FUTURE    TECHNOL   INDUSTRY
            --------   --------   --------   --------   --------   --------
  COMFORT      1.000
  ENVIRON      0.099      1.000
     WORK      0.201     -0.083      1.000
   FUTURE      0.346     -0.028      0.479      1.000
  TECHNOL      0.090      0.464     -0.104     -0.036      1.000
INDUSTRY       0.182      0.411     -0.008      0.103      0.435      1.000
  BENEFIT      0.408     -0.037      0.209      0.377     -0.014      0.118

        Correlation Matrix

             BENEFIT
            --------
  BENEFIT      1.000
```

Using DWLS, it took 11 iterations to obtain the following results.

```
LAMBDA-X

            Positive   Negative
            --------   --------
  COMFORT      0.528      0.197
              (0.074)    (0.076)
                7.167      2.595

  ENVIRON       - -       0.661
                         (0.069)
                           9.609

     WORK       0.536       - -
              (0.064)
                8.310

   FUTURE       0.768       - -
              (0.061)
               12.588

  TECHNOL       - -       0.692
                         (0.061)
                          11.365

 INDUSTRY       - -       0.631
                         (0.067)
                           9.417

  BENEFIT      0.535       - -
              (0.064)
                8.322
```

PHI

	Positive	Negative
Positive	1.000	
Negative	-0.025	1.000
	(0.093)	
	-0.264	

All factor loadings are statistically significant, but the estimated factor correlation is very small and not significant. The estimated variances of the δ_i are

THETA-DELTA

COMFORT	ENVIRON	WORK	FUTURE	TECHNOL	INDUSTRY
0.687	0.564	0.713	0.410	0.521	0.602
(0.127)	(0.136)	(0.122)	(0.138)	(0.132)	(0.132)
5.409	4.148	5.825	2.974	3.957	4.566

THETA-DELTA

BENEFIT
0.713
(0.122)
5.834

The following goodness of fit statistics are provided when DWLS estimation is used. From these, we conclude that the model does not describe the data well.

Goodness-of-Fit Statistics

Degrees of Freedom for (C1)-(C3),C(5)	12
Maximum Likelihood Ratio Chi-Square (C1)	47.235 (P = 0.0000)
Browne's (1984) ADF Chi-Square (C2_NT)	48.780 (P = 0.0000)
Browne's (1984) ADF Chi-Square (C2_NNT)	20.009 (P = 0.0669)
Satorra-Bentler (1988) Scaled Chi-Square (C3)	24.655 (P = 0.0165)
Satorra-Bentler (1988) Adjusted Chi-Square (C4)	21.719 (P = 0.0219)
Degrees of Freedom for C4	10.571
Chi-Square Scaled and Shifted (C5)	23.878 (P = 0.0211)
P-Value of C1 under Non-Normality	= 0.0231