Confirmatory factor analysis of ordinal data without missing values

Contents

1. Introduction 1

2. Estimation using adaptive quadrature 2

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.

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.

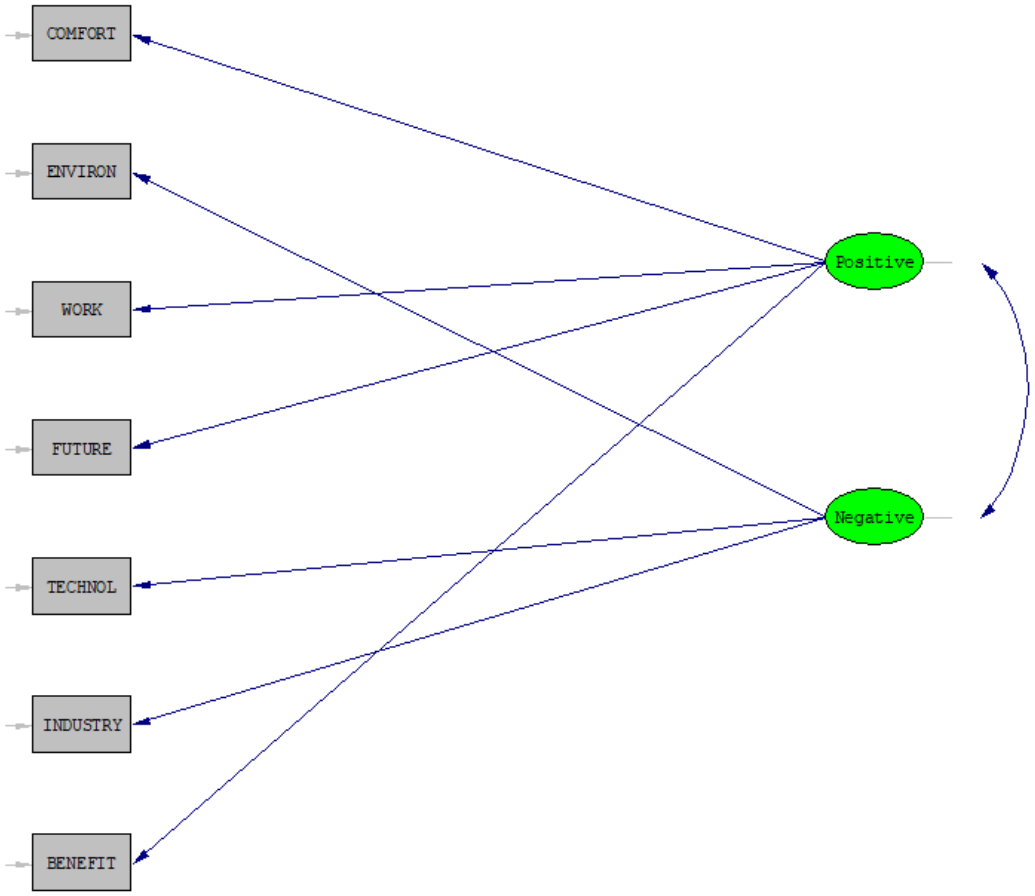
SCITECH.LSF

	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 adaptive quadrature

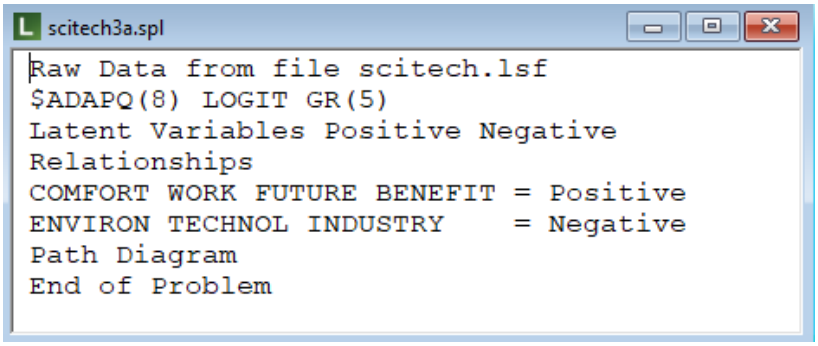
The use of adaptive quadrature estimation in CFA and other models are discussed in various sources, such as Schilling and Bock (2005), Bryant and Jöreskog (2016) and Vasdekis, Cagnone and Moustaki (2012).

Based on previous results, we wish to fit the following confirmatory factor analysis to the data.



The syntax to fit this model is given below. Four link functions are available, these being the logit, the probit, the log-log and the complimentary log-log. Here a logit link function is used, and 8 quadrature points are specified. The equivalent LISREL syntax for this model is given in **scitech3b.lis**. As a general rule, one should:

- use 8 quadrature points for models with one latent variable
- 5 to 10 for models with two or three latent variables,
- 5 to 6 for models with four or five latent variables, and
- 3 to 4 for models with six to ten latent variables.



Fitting information reported from the output file (see below) does not provide a chi-square goodness of fit value, as there is no appropriate saturated model available to compare with.

```
    Number of quadrature points =              8
    Number of free parameters =               29
    Number of iterations used =                8

    -2lnL (deviance statistic) =       5841.79608
    Akaike Information Criterion       5899.79608
    Schwarz Criterion                  6014.96268
```

From the LISREL output, we obtain the following estimates:

```
        LAMBDA-X

           Positive   Negative
           --------   --------
 COMFORT      1.046       - -
             (0.190)
               5.503

 ENVIRON       - -       1.622
                        (0.250)
                          6.490

    WORK      1.221       - -
             (0.182)
               6.715

  FUTURE      2.289       - -
             (0.484)
               4.725

 TECHNOL       - -       1.744
                        (0.279)
                          6.247

INDUSTRY       - -       1.530
                        (0.242)
                          6.316

 BENEFIT      1.095       - -
             (0.183)
               5.970
```

PHI

	Positive	Negative
Positive	1.000	
Negative	0.020	1.000
	(0.081)	
	0.250	

All the factor loadings are statistically significant, but the correlation between the two factors is not. This implies that we can assume that the factors are uncorrelated.

Since ordinal variables have no scale or unit of measurement, this model does not have any error terms. Factor loadings and threshold parameters are unstandardized. As a result, the program reports estimated threshold parameters instead of a THETA-DELTA matrix.

Threshold estimates and standard deviations

Threshold	Estimates	S.E.	Est./S.E.
TH1_COMFORT	-4.86789	0.49163	-9.90162
TH2_COMFORT	-2.64287	0.22365	-11.81713
TH3_COMFORT	1.46795	0.15931	9.21422
TH1_ENVIRON	-3.43055	0.31632	-10.84525
TH2_ENVIRON	-1.25484	0.17941	-6.99409
TH3_ENVIRON	1.00321	0.16769	5.98253
TH1_WORK	-2.92080	0.23907	-12.21714
TH2_WORK	-0.90018	0.14270	-6.30835
TH3_WORK	2.26337	0.20288	11.15644
TH1_FUTURE	-5.23237	0.73195	-7.14854
TH2_FUTURE	-2.21286	0.35796	-6.18195
TH3_FUTURE	1.96282	0.32332	6.07088
TH1_TECHNOL	-4.16898	0.40088	-10.39964
TH2_TECHNOL	-1.49502	0.20337	-7.35119
TH3_TECHNOL	1.09062	0.18085	6.03045
TH1_INDUSTRY	-4.58387	0.42731	-10.72728
TH2_INDUSTRY	-2.42542	0.24091	-10.06765
TH3_INDUSTRY	0.45819	0.14732	3.11025
TH1_BENEFIT	-3.34883	0.27681	-12.09782
TH2_BENEFIT	-0.99225	0.14060	-7.05739
TH3_BENEFIT	1.68846	0.16873	10.00674

Recall that each of the ordinal variables considered here had 4 possible outcomes, *i.e.*, 4 categories. As a result, three thresholds are reported for each of the variables. Note the difference in the range of threshold estimates of the first variable, COMFORT and that of the second to last variable INDUSTRY.