Confirmatory factor analysis of ordinal data without missing values

Contents

1. Introduction

This example is based on six political efficacy measurements as described in Aish & Jöreskog (1990). The dataset **EFFICACY.LSF** consists of 1719 cases obtained in a USA sample where the number -999999.0 denotes a missing value.

efficacy.lsf

	NOSAY	VOTING	COMPLEX	NOCARE	TOUCH	INTEREST
1	2.00	2.00	1.00	1.00	1.00	1.00
2	2.00	3.00	3.00	3.00	2.00	3.00
3	3.00	2.00	2.00	3.00	3.00	3.00
4	3.00	3.00	2.00	3.00	2.00	3.00
5	2.00	2.00	1.00	2.00	2.00	2.00
6	2.00	2.00	1.00	1.00	2.00	1.00
7	3.00	2.00	2.00	3.00	3.00	3.00
8	2.00	2.00	2.00	2.00	1.00	2.00
9	3.00	1.00	2.00	2.00	2.00	2.00
10	2.00	1.00	2.00	2.00	1.00	1.00
11	2.00	2.00	1.00	2.00	2.00	2.00
12	1.00	1.00	1.00	1.00	2.00	-999999.00
13	1.00	-999999.00	1.00	1.00	1.00	1.00
14	3.00	3.00	2.00	3.00	3.00	3.00

The data are the responses to the following statements:

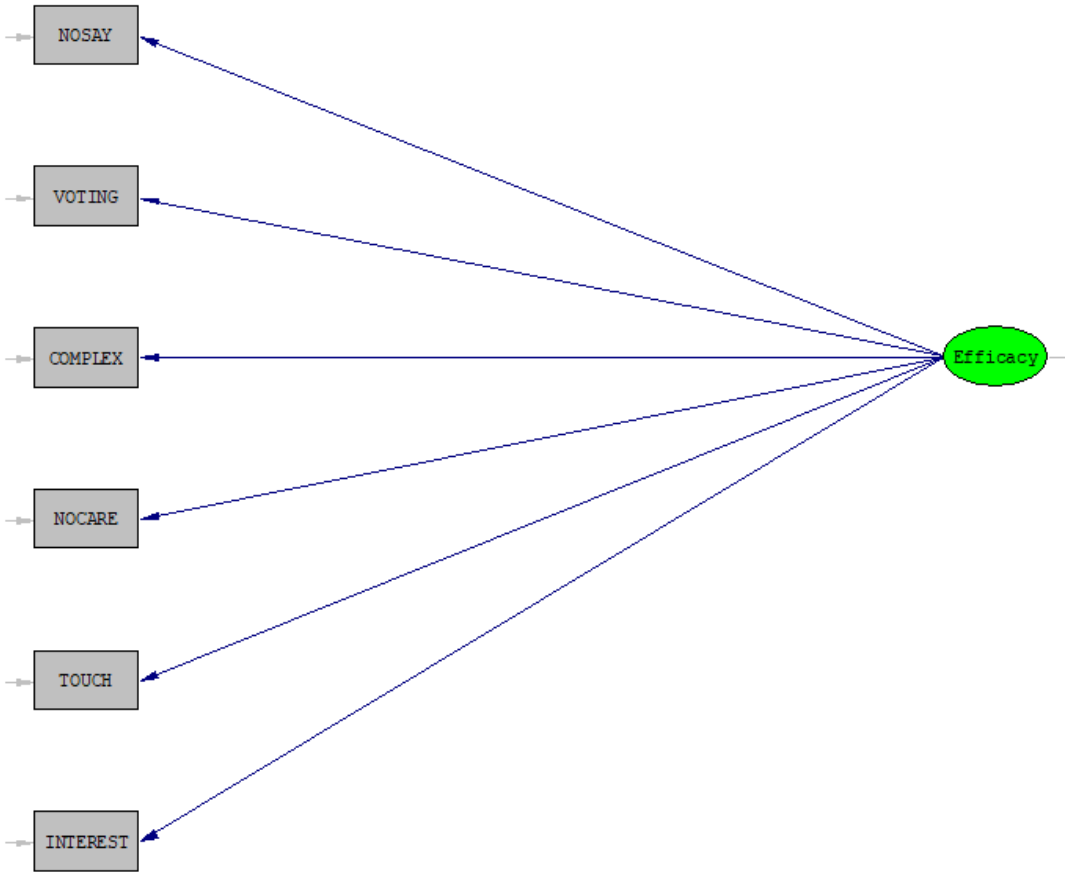
- NOSAY: "People like me have no say in what the government does."
- VOTING: "Voting is the only way that people like me can have any say about how the government runs things."

- COMPLEX: "Sometimes politics and government seem so complicated that a person like me cannot really understand what is going on."
- NOCARE: "I don't think that public officials care much about what people like me think."
- TOUCH: "Generally speaking, those who elect to Congress in Washington lose touch with the people pretty quickly."
- INTEREST: "Parties are only interested in people's votes but not in their opinions."

The ordered categories are:

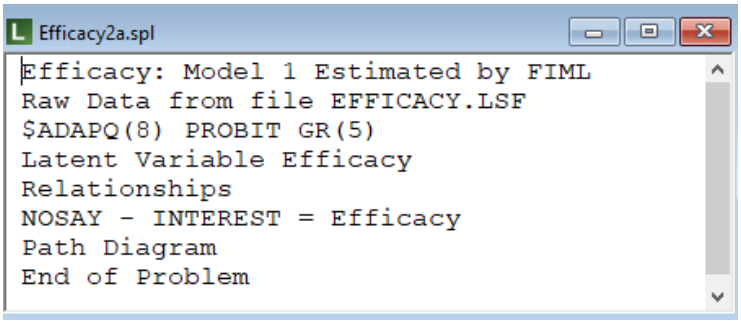
- 1: agree strongly
- 2: agree
- 3: disagree
- 4: disagree strongly

It is hypothesized that the six variables are indicators of a single unidimensional latent variables, as shown in the conceptual path diagram below.



2. Estimation using adaptive quadrature – one factor

To fit the model proposed above using FIML, we use the following syntax (**efficacy2a.spl**), requesting 8 quadrature points and the use of the probit link function.



The maximum likelihood estimates for this analysis are as follows:

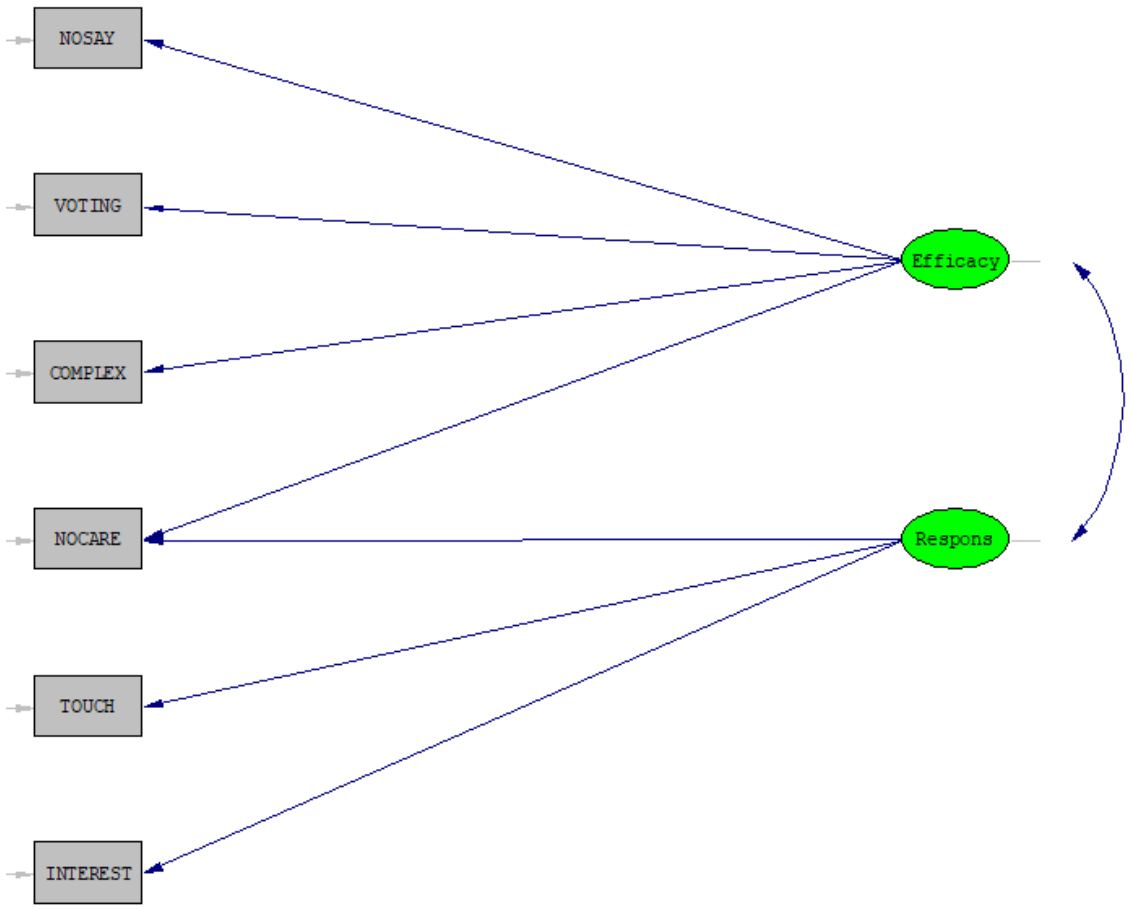
```
LISREL Estimates (Maximum Likelihood)

         Measurement Equations


    NOSAY = 0.739*Efficacy, Errorvar.= 1.000, R² = 0.353
 Standerr  (0.0407)
 Z-values   18.154
 P-values   0.000

   VOTING = 0.377*Efficacy, Errorvar.= 1.000, R² = 0.124
 Standerr  (0.0324)
 Z-values   11.643
 P-values   0.000

  COMPLEX = 0.601*Efficacy, Errorvar.= 1.000, R² = 0.265
 Standerr  (0.0375)
 Z-values   16.042
 P-values   0.000

   NOCARE = 1.656*Efficacy, Errorvar.= 1.000, R² = 0.733
 Standerr  (0.103)
 Z-values   16.008
 P-values   0.000

    TOUCH = 1.185*Efficacy, Errorvar.= 1.000, R² = 0.584
 Standerr  (0.0632)
 Z-values   18.754
 P-values   0.000

 INTEREST = 1.361*Efficacy, Errorvar.= 1.000, R² = 0.649
 Standerr  (0.0744)
 Z-values   18.292
 P-values   0.000
```

The smallest loading is associated with the variable VOTING. The output also gives the value of some fit statistics and the value of the deviance statistic at convergence. However, as we have nothing to compare the latter with, it is not particularly informative regarding model fit.

```
   -2lnL (deviance statistic) =      19934.56515
     Akaike Information Criterion      19982.56515
     Schwarz Criterion                 20113.22712
```

3. Estimation using adaptive quadrature – two factors

It has been hypothesized in the political science literature that the six variables are really indicators of two factors, where Factor 1 represents *Internal Efficacy* measured by NOSAY, VOTING and COMPLEX. Factor 2 represents *External Efficacy* measured by TOUCH and INTEREST. We denote the two factors by Efficacy and Respons. The statement NOCARE, contains two referents: "public officials" and "people like me". As such, it can be argued that it may measure both factors. Here we opt to treat NOCARE as a complex variable measuring both Efficacy and Respons as shown in the conceptual path diagram below.



We use SIMPLIS syntax to fit this model to the data:

Efficacy3a.spl

```
Efficacy: Model 2 Estimated by FIML
Raw Data from file EFFICACY.LSF
$ADAPQ(8) PROBIT GR(3)
Latent Variables Efficacy Respons
Relationships
NOSAY - NOCARE = Efficacy
NOCARE - INTEREST = Respons
Path Diagram
End of Problem
```

For this model, we obtain the following output:

```
     Measurement Equations


    NOSAY = 0.916*Efficacy, Errorvar.= 1.000, R² = 0.456
 Standerr  (0.0601)
 Z-values   15.245
 P-values   0.000

   VOTING = 0.461*Efficacy, Errorvar.= 1.000, R² = 0.175
 Standerr  (0.0385)
 Z-values   11.973
 P-values   0.000

  COMPLEX = 0.686*Efficacy, Errorvar.= 1.000, R² = 0.320
 Standerr  (0.0455)
 Z-values   15.093
 P-values   0.000

   NOCARE = 0.824*Efficacy + 0.901*Respons, Errorvar.= 1.000, R² = 0.723
 Standerr  (0.131)          (0.109)
 Z-values   6.268            8.301
 P-values   0.000            0.000

    TOUCH = 1.333*Respons, Errorvar.= 1.000, R² = 0.640
 Standerr  (0.0778)
 Z-values   17.138
 P-values   0.000

 INTEREST = 1.607*Respons, Errorvar.= 1.000, R² = 0.721
 Standerr  (0.107)
 Z-values   15.000
 P-values   0.000


         Correlation Matrix of Independent Variables

            Efficacy    Respons
            --------   --------
 Efficacy      1.000

  Respons      0.752      1.000
             (0.030)
              25.017
```

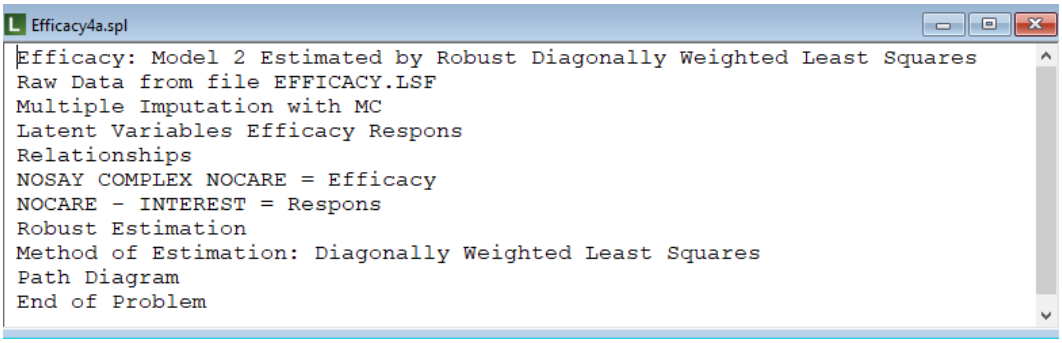
We note that the relationship between NOCARE and Efficacy is about just as strong as between NOCARE and Respons, as indicated by the estimated factor loadings of 0.824 and 0.901 respectively. The correlation matrix of independent variables indicate a significant correlation between the two factors.

```
     Number of quadrature points =              8
     Number of free parameters =               26
     Number of iterations used =               16

     -2lnL (deviance statistic) =      19858.05634
     Akaike Information Criterion      19910.05634
     Schwarz Criterion                 20051.60681
```

This model has two more parameters than the previous model, as we allowed NOCARE to associate with both factors, and we added a factor in the current model. The difference in deviances between the current and previous model can thus be used as a chi-square statistic with 2 degrees of freedom to evaluate whether the additional two parameters made a significant contribution to describing the data. The chi-square is 76.507, which is significant and we conclude that the second model fits better than the first.

1. Estimation using DWLS – two factors

We now repeat the fitting of the second model, using DWLS instead of adaptive quadrature. To do so, we use the syntax file **efficacy4a.spl** (LISREL syntax **efficacy4b.lis**).



Output obtained for this run is shown below. These results are somewhat different from those obtained by adaptive quadrature. It should be kept in mind that the two approaches estimate the same model, but they use methods based on different assumptions.

The number of estimated parameters may be the same, but the parameters are not directly comparable. To compare them, one has to obtain the completely standardized solution that can be obtained as a by-product in the FIML approach and the DWLS solution which is already standardized.

```
LISREL Estimates (Robust Diagonally Weighted Least Squares)

         LAMBDA-X

            Efficacy    Respons
            --------   --------
    NOSAY      0.583       - -
              (0.035)
               16.491

  COMPLEX      0.332       - -
              (0.021)
               15.954

   NOCARE      0.608      0.148
              (0.160)    (0.161)
                3.801      0.918

    TOUCH       - -       0.526
                         (0.014)
                          37.652

 INTEREST       - -       0.641
                         (0.016)
                          39.222
```

PHI

	Efficacy	Respons
Efficacy	1.000	
Respons	0.841	1.000
	(0.046)	
	18.329	

THETA-DELTA

NOSAY	COMPLEX	NOCARE	TOUCH	INTEREST
0.557	0.289	0.135	0.154	0.144
(0.060)	(0.024)	(0.047)	(0.026)	(0.034)
9.298	12.178	2.861	6.043	4.213

Squared Multiple Correlations for X - Variables

NOSAY	COMPLEX	NOCARE	TOUCH	INTEREST
0.379	0.276	0.801	0.642	0.741

Goodness-of-Fit Statistics

Degrees of Freedom for (C1)-(C3),C(5)	3
Maximum Likelihood Ratio Chi-Square (C1)	7.649 (P = 0.0539)
Browne's (1984) ADF Chi-Square (C2_NT)	6.969 (P = 0.0729)
Browne's (1984) ADF Chi-Square (C2_NNT)	2.540 (P = 0.4681)
Satorra-Bentler (1988) Scaled Chi-Square (C3)	2.802 (P = 0.4232)
Satorra-Bentler (1988) Adjusted Chi-Square (C4)	2.791 (P = 0.4229)
Degrees of Freedom for C4	2.988
Chi-Square Scaled and Shifted (C5)	2.802 (P = 0.4231)
P-Value of C1 under Non-Normality	= 0.4228