Cross-sectional ordinal data

Contents

1   Introduction

This example is the second of a set of examples extracted from a note by K.G. Jöreskog first posted on the SSI website in 2005 with the title "*Structural Equation Modeling with Ordinal Variables using LISREL*".

In the previous example, I introduced a small data set of six variables from the cross-sectional USA sample of the Political Action Survey. After imputation of some missing values and listwise deletion after imputation I ended up with 1589 cases with complete data on all six variables.

Political scientists assume that these items measure a uni-dimensional trait called political efficacy which has been defined as the feeling that individual political action does have, or can have, an impact upon the political process (Campbell, et al., 1954). In this section, I will demonstrate how the assumption of uni-dimensionality can be tested. I will also describe the statistical model used in PRELIS and LISREL for this purpose and discuss the assumptions of this model.

2 Ordinal Variables and Underlying Variables

Observations on an ordinal variable represent responses to a set of ordered categories, such as a four- category Likert scale. It is only assumed that a person who selected a specific category has more of a characteristic than if he/she had chosen a lower category, but we do not know how much more. Ordinal variables are not continuous variables and should not be treated as if they are. It is common practice to treat scores 1, 2, 3, ... assigned to categories as if they have metric properties but this is wrong. Ordinal variables do not have origins or units of

measurements. Means, variances, and covariances of ordinal variables have no meaning. The only information we have are counts of cases in each cell of a multiway contingency table. To use ordinal variables in structural equation models requires other techniques than those that are traditionally employed with continuous variables. For each ordinal variable z (which may be a y- or a x-variable in LISREL sense), it is assumed that there is an underlying continuous variable z^*. This continuous variable z^* represent the attitude underlying the ordered responses to z and is assumed to have a range from $-\infty$ to $+\infty$. The underlying variable z^* can be used in structural equation modeling, not the observed variable z. The underlying variable assigns a metric to the ordinal variable.

If z has m categories labeled 1, 2, ..., m, the connection between z and z^* is

$$z = i \Leftrightarrow \tau_{i-1} < z^* < \tau_i, \quad i = 1, 2, ..., m,$$

where

$$-\infty = \tau_0 < \tau_1 < \tau_2 < ... < \tau_{m-1} < \tau_m = +\infty,$$

are parameters called threshold values. With m categories, there are m - 1 threshold parameters $\tau_1 < \tau_2 < ... < \tau_{m-1}$.

Because we have only ordinal information, the distribution of z^* is determined only up to a monotonic transformation. In principle, one can choose any continuous distribution for z^*. However, any continuous variable with a density and a distribution function can be transformed by a monotonic transformation to a normal distribution. It is therefore convenient to choose the standard normal distribution with density function $\phi(u)$ and distribution function $\Phi(u)$ for z^*. Then the probability of a response in category i is

$$\pi_i = \Phi^{-1} \Pr[z = i] = \Pr[\tau_{i-1} < z^* < \tau_i] = \int_{\tau_{i-1}}^{\tau_i} \phi(u) du = \Phi(\tau_i) - \Phi(\tau_{i-1}), \qquad (1)$$

so that

$$\tau_i = \Phi^{-1}(\pi_1 + \pi_2 + ... + \pi_i), \quad i = 1, ..., m-1, \qquad (2)$$

where Φ^{-1} is the inverse of the standard normal distribution function. The quantity $(\pi_1 + \pi_2 + ... + \pi_i)$ is the probability of a response in category i or lower.

The probabilities π_i are unknown population quantities. In practice, π_i can be estimated consistently by the corresponding percentage p_i of responses in category i which is given in the PRELIS output. Then, estimates of the thresholds can be obtained as

$$\hat{\tau}_i = \Phi^{-1}(p_1 + p_2 + ... + p_i), \quad i = 1, ..., m-1. \qquad (3)$$

The quantity $\left(p_1 + p_2 + ... + p_i \right)$ is the proportion of cases in the sample responding in category i or lower. Equation (3) is in fact the maximum likelihood estimator of τ_i based on the univariate marginal sample data. But the model is saturated; there are m - 1 parameters τ_i and there are m - 1 independent sample proportions p_i. The fit is perfect since

$$\hat{\pi}_i = \Phi\left(\hat{\tau}_i \right) - \Phi\left(\hat{\tau}_{i-1} \right) = p_i.$$

The estimation of thresholds is illustrated in Fig. 1. Suppose there are 8, 41, 39, and 12% responding in category 1, 2, 3, and 4, respectively. Cumulatively this is 8, 49, 88, 100%. The first threshold is located where the area under the normal to the left of the threshold is 8%. The second threshold is located where the area under the normal to the left of the threshold is 49%. The third threshold is located where the area under the normal to the left of the threshold is 88%. This gives approximately

$$\hat{\tau}_1 = -1.404, \quad \hat{\tau}_2 = -0.025, \quad \hat{\tau}_3 = 1.075.$$

| 8% | 41% | 39% | 12% |
| 8% | 49% | 88% | 100% |
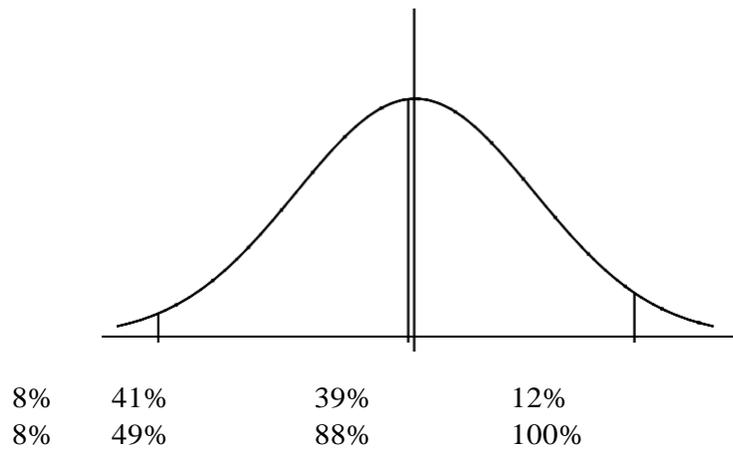
Figure 1: Illustrating Thresholds

3 Polychoric Correlations

Let z_1 and z_2 be two ordinal variables with m_1 and m_2 categories, respectively. Their marginal distribution in the sample is represented by a contingency table

$$\begin{pmatrix} n_{11} & n_{12} & \cdots & n_{1m_2} \\ n_{21} & n_{22} & \cdots & n_{2m_2} \\ \vdots & \vdots & \ddots & \vdots \\ n_{m_1 1} & n_{m_1 2} & \cdots & n_{m_1 m_2} \end{pmatrix},$$

where n_{ij} is the number of cases in category i on variable 1 and in category j on variable 2. Since the underlying variables z_1^* and z_2^* are normal with zero means, unit variances, it is natural to assume that z_1^* and z_2^* are standard bivariate normal with correlation ρ. However, unlike the univariate marginal normality of z_1^* and z_2^*, which

cannot be falsified using univariate marginal data, the bivariate normality of z_1^* and z_2^* is an assumption which is testable using bivariate marginal data. Normality of z_1^* and z_2^* does not imply bivariate normality of z_1^* and z_2^*.

The *polychoric correlation* is the correlation ρ in the bivariate normal distribution of the underlying variables z_1^* and z_2^*. If $m_1 = m_2 = 2$ this is called a *tetrachoric correlation*.

Let $\tau_1^{(1)}, \tau_2^{(1)}, ..., \tau_{m-1}^{(1)}$ be the thresholds for variable z_1^* and let $\tau_1^{(2)}, \tau_2^{(2)}, ..., \tau_{m-1}^{(2)}$ be the thresholds for variable z_2^*. Then the polychoric correlation can be estimated by maximizing the log-likelihood of the multinomial distribution,

$$\ln L = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} n_{ij} \log \pi_{ij}(\boldsymbol{\theta}),$$

where

$$\pi_{ij}(\boldsymbol{\theta}) = \Pr[z_1 = i, z_2 = j] = \int_{\tau_{i-1}^{(1)}}^{\tau_i^{(1)}} \int_{\tau_{i-1}^{(2)}}^{\tau_i^{(2)}} \phi_2(u,v) du dv, \qquad (4)$$

and

$$\phi_2(u,v) = \frac{1}{2\pi \sqrt{(1-\rho^2)}} e^{-\frac{1}{2(1-\rho^2)}(u^2 - 2\rho uv + v^2)},$$

is the standard bivariate normal density with correlation ρ.

The model defined by (4) expresses the $m_1 m_2$ probabilities $\pi_{ij}(\boldsymbol{\theta})$ as functions of the parameter vector

$$\boldsymbol{\theta} = \left(\tau_1^{(1)}, \tau_2^{(1)}, ..., \tau_{m_1-1}^{(1)}, \tau_1^{(2)}, \tau_2^{(2)}, ..., \tau_{m_2-1}^{(2)}, \rho \right), \qquad (5)$$

consisting of the thresholds for the two variables and the polychoric correlation ρ. Maximizing $\ln L$ is equivalent to minimizing the fit function

$$F(\boldsymbol{\theta}) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} p_{ij}[\ln p_{ij} - \ln \pi_{ij}(\boldsymbol{\theta})] = \sum_{ij} p_{ij}[p_{ij} / \pi_{ij}(\boldsymbol{\theta})], \qquad (6)$$

where $p_{ij} = n_{ij} / N$ are the sample proportions.

The estimation of the polychoric correlation is illustrated in Fig. 2. Here the numbers are the sample proportions and the vertical and horizontal lines represent the thresholds. The ellipse represents the standard bivariate normal distribution with correlation ρ. The value of ρ determines the orientation and concentration of the ellipse. This can be fitted to the sample proportions.
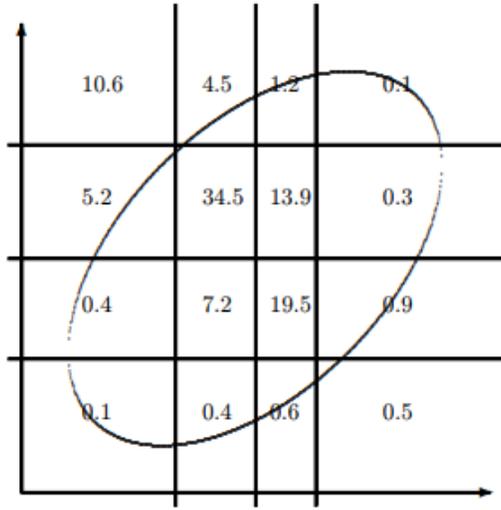
Figure 2: Estimating the Polychoric Correlation

PRELIS estimates the parameters by a two-step procedure, see Olsson (1979). In the first step, the thresholds are estimated from the univariate marginal distributions by (3). In the second step, the polychoric correlations are estimated from the bivariate marginal distributions by minimizing (6) for given thresholds. The parameters can also be estimated by a one-step procedure. It is no problem to estimate the thresholds and the polychoric correlations simultaneously, but this is not necessary and it is not practical. It is not necessary because the estimates are almost the same as with the two-step procedure. It is not practical because it would yield different thresholds for the same variable when paired with different variables.

4   Testing Underlying Bivariate Normality

To test the model, one can use the likelihood ratio (LR) test statistic

$$\chi^2_{LR} = 2\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} n_{ij}\ln(p_{ij}/\hat{\pi}_{ij}) = 2N\sum_{i=1}^{m_1}\sum_{j=1}^{m_2} p_{ij}\ln(p_{ij}/\hat{\pi}_{ij}) = 2NF(\hat{\theta}), \qquad (7)$$

where $\hat{\theta}$ is the estimated parameter vector and $\hat{\pi}_{ij} = \pi_{ij}\left(\hat{\theta}\right)$. Hence, this χ^2 is just $2N$ times the minimum value of the fit function (6). If the model holds, this is distributed approximately as χ^2 with degrees of freedom

$$d = \left(m_1 m_2 - 1\right) - (m_1 - 1) - (m_2 - 1) - 1 = m_1 m_2 - m_1 - m_2. \qquad (8)$$

Alternatively, one can use the goodness-of-fit (GF) test statistic

$$\chi^2_{GF} = 2\sum_{i=1}^{m_1}[\sum_{j=1}^{m_2}(n_{ij} - N\hat{\pi}_{ij})^2/(N\hat{\pi}_{ij})] = N\sum_{i=1}^{m_1}\sum_{j=1}^{m_2}(p_{ij} - \hat{\pi}_{ij})^2/\hat{\pi}_{ij}. \qquad (9)$$

If the model holds, both statistics (7) and (9) have the same asymptotic distribution. In practice, when the fit is good the LR and GF statistic are quite close but when the model does not fit well they can be quite different.

If $m_1 = m_2 = 2$, $d = 0$. That is, if both variables are dichotomous, the degrees of freedom is zero, the model is saturated and it is not possible to test underlying bivariate normality.

Because of the two-step procedure used in PRELIS, the χ^2_{LR} reported by PRELIS slightly over-estimates the correct asymptotic chi-square. However, this is of no practical importance since the parameter estimates are essentially the same, whether the two-step or the one-step procedure is used. Table 3 gives an example based on
the data for NOSAY vs VOTING.

Table 3: Two-Step vs One-Step Procedure

	Two-step method			One-step method		
Thresholds for NOSAY	-1.267	-0.222	1.423	-1.267	-0.222	1.419
Thresholds for VOTING	-0.942	0.220	1.666	-0.942	0.218	1.666
Polychoric correlation	0.329			0.330		
LR Chi-square with 8 df	222.381			222.365		
GF Chi-square with 8 df	325.120			324.804		

This example shows that the parameter estimates and chi-squares are essentially the same for the two methods. It also shows that the LR χ^2 and the GF χ^2 can be quite different when the model does not fit. Lack of fit will be discussed in the next section. The LR and GF statistics are sometimes denoted G2 and X2, respectively.

5 Estimating the Polychoric Correlations and their Asymptotic Covariance Matrix

We now return to the analysis of the data on the six political efficacy items. Do the six items measure one uni-dimensional latent variable? For this to work there must be a clear correspondence between the values of the latent variable and the categories of the ordinal variables, see Costner (1969). In this case, people who are high on Efficacy are supposed to disagree or disagree strongly and people who are low on Efficacy should agree or agree strongly to these items. If this is the case, there would be a positive association between the latent variable Efficacy and each ordinal variable. But isn't something wrong with VOTING? If I am high on Efficacy and I believe that voting is the only way I can influence politics, then I would agree or agree strongly to the VOTING statement. This fact in itself is sufficient to suggest that the VOTING item should be excluded from further consideration. However, to begin with, I shall keep it to demonstrate what consequences this has in the data analysis.

To estimate the one-factor model there are two steps:

- Use PRELIS to estimate the polychoric correlations and their asymptotic covariance matrix.

- Use these matrices in LISREL to estimate the one-factor model with weighted least squares (WLS).

The PRELIS step is described in this section and the LISREL step in Section 5. For the PRELIS step, use the following syntax file (file **ORD21.PRL**):

```
EFFICACY: PRELIS Run 4
Computing Polychoric Correlations and Asymptotic Covariance Matrix
Data Ninputvariables = 6
Labels
NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
RA=EFFICACY.IMP
CLabels NOSAY - INTEREST 1=AS 2=A 3=D 4=DS
Output BT MA=PM PM=EFFICACY.PM AC=EFFICACY.ACP
```

The data is in the file **EFFICACY.IMP**. MA = PM on the Output line means that the matrix of polychoric correlations should be estimated, PM = EFFICACY.PM means that this matrix will be saved in the file **EFFICACY.PM** and AC = EFFICACY.ACP means that the asymptotic covariance matrix will be saved in the file **EFFICACY.ACP**. The first file is in text (ASCII, readable) format and the second is in binary (unreadable) format.
If you have saved the imputed data as a LSF file, **EFFICACY.LSF**, say, instead of as the text file **EFFICACY.IMP**, you can use a much shorter PRELIS input file4 (file **ORD21A.PRL**):

```
EFFICACY: PRELIS Run 4
Computing Polychoric Correlations and Asymptotic Covariance Matrix
SY=EFFICACY.LSF
Output BT MA=PM PM=EFFICACY.PM AC=EFFICACY.ACP
```

An easy way to obtain the file **EFFICACY.LSF** is to add RA = EFFICACY.LSF on the Output line in PRELIS Run 3, see file **ORD13A.PRL**.

The program will write a separate output file listing all response patterns occurring in the sample and the frequency of occurrence of each pattern. This file has the name **inputfilename.FREQ**. For example, if the input file for PRELIS Run 4 is **ORD21.PRL**, the FREQ file will be **ORD21.FREQ**. The 20 most common response patterns will also be listed in the ordinary output file. The FREQ file should be regarded as a data file. It gives the data in the most concise form. The FREQ file may be read by PRELIS by specifying the frequency variable as a weight variable. The reason for the FREQ file having the name of the input file rather than the name of the data file is that it depends on the PRELIS commands in the input file.

In the Efficacy example, each variable has four categories. Thus, there are 4096 possible response patterns but since we have data on only 1589 cases, every response pattern cannot be present in the data. In fact, there are only 476 different response patterns occurring in the sample. All these are given in the FREQ file. The output file lists the 20 most common response patterns as

There are 476 distinct response patterns, see FREQ-file. The 20 most common patterns are:

```
106        A    A    A    A    A    A
 70        D    D    A    D    D    D
 51        D    A    A    A    A    A
 45        D    D    D    D    D    D
 45        D    D    A    A    A    A
 40        D    A    A    D    D    D
 32        D    D    A    D    A    A
 31        D    D    A    D    A    D
 25        A    A    AS   A    A    A
 23        AS   AS   AS   AS   AS   AS
 20        D    A    A    D    A    A
 19        A    A    AS   AS   AS   AS
 19        A    D    A    A    A    A
 18        D    D    A    A    A    D
 16        D    D    D    D    A    D
 16        D    A    A    D    A    D
 15        D    D    A    D    D    A
 15        D    A    D    D    D    D
 15        A    AS   A    A    A    A
 14        A    A    A    D    A    A
```

It is seen that the most common response pattern is to answer *Agree* to all six items (106 cases). The second most common response pattern is to answer *Agree* to COMPLEX and *Disagree* to all the other five items (70 cases). To investigate whether there is a tendency to give the same response to all six items, one can note that there are 45 cases who *Disagree* to all six items and 23 cases who *Agree Strongly* to all six items. Further screening of the FREQ file reveals that there are only 3 cases who *Disagree Strongly* to all six items. Note that among the 20 most common response patterns there are no response *Disagree Strongly* on any of the six variables.

The output gives the following table of univariate marginal parameters. This refers to the underlying variables. An alternative parameterization is presented in Section 7.

```
Total Sample Size(N) =    1554

 Univariate Marginal Parameters

 Variable     Mean St. Dev.    Thresholds
 --------     ---- --------    ----------
    NOSAY     0.000    1.000  -1.265   -0.238    1.428
   VOTING     0.000    1.000  -0.937    0.210    1.682
  COMPLEX     0.000    1.000  -0.846    0.725    1.767
   NOCARE     0.000    1.000  -1.023    0.155    1.807
    TOUCH     0.000    1.000  -0.973    0.527    2.142
 INTEREST     0.000    1.000  -0.978    0.323    2.082
```

With 6 variables there are 15 different pairs of variables. The contingency tables for each of these is given in the output as

Bivariate Distributions for Ordinal Variables (Frequencies)

	VOTING				COMPLEX				NOCARE			
NOSAY	AS	A	D	DS	AS	A	D	DS	AS	A	D	DS
AS	69	51	27	13	83	52	16	9	80	58	19	3
A	80	297	85	9	120	287	55	9	104	278	88	1
D	92	275	413	24	89	491	201	23	44	281	455	24
DS	30	12	51	26	17	51	32	19	10	18	64	27

	TOUCH				INTEREST			
NOSAY	AS	A	D	DS	AS	A	D	DS
AS	75	67	16	2	84	50	24	2
A	99	311	60	1	99	281	91	0
D	62	411	316	15	57	357	378	12
DS	21	43	48	7	15	31	58	15

	COMPLEX				NOCARE				TOUCH			
VOTING	AS	A	D	DS	AS	A	D	DS	AS	A	D	DS
AS	102	125	37	7	73	108	78	12	82	134	49	6
A	131	389	99	16	97	327	203	8	100	398	133	4
D	63	341	151	21	55	190	311	20	58	275	233	10
DS	13	26	17	16	13	10	34	15	17	25	25	5

	INTEREST			
VOTING	AS	A	D	DS
AS	84	104	74	9
A	105	351	177	2
D	50	246	269	11
DS	16	18	31	7

	NOCARE				TOUCH				INTEREST			
COMPLEX	AS	A	D	DS	AS	A	D	DS	AS	A	D	DS
AS	133	115	56	5	131	142	31	5	122	141	41	5
A	84	418	363	16	90	521	263	7	101	444	326	10
D	16	88	184	16	23	151	124	6	24	118	159	3
DS	5	14	23	18	13	18	22	7	8	16	25	11

	TOUCH				INTEREST			
NOCARE	AS	A	D	DS	AS	A	D	DS
AS	150	73	13	2	154	70	14	0
A	84	466	83	2	79	437	117	2
D	20	280	315	11	17	200	398	11
DS	3	13	29	10	5	12	22	16

```
                INTEREST
           ------------------
   TOUCH   AS    A    D    DS
           ------------------
    AS     167   70   19    1
    A       80  531  216    5
    D        7  112  306   15
    DS       1    6   10    8
```

Note that there are two zero cells in these tables, one in the table for NOSAY and INTEREST and one in the table for NOCARE and INTEREST. The fit function (6) is not defined if ρ_{ij} is zero. In the estimation of the polychoric correlations, PRELIS skips such zero cells. But too many zero cells in a table can be problematic and give estimates that are imprecise and unreliable. If there is only one nonzero cell in a row or a column, the estimation procedure breaks down. A zero cell is particularly problematic in the case when both variables are dichotomous because then the tetrachoric correlation is undefined. In this case PRELIS replaces the zero by a half. But the estimate of the tetrachoric correlation is very sensitive to changes in the value used to substitute the zero cell. Other programs use different values and therefore get different estimates of the tetrachoric correlation.

From the table of NOSAY vs VOTING it is seen that of the 160 cases who agree strongly with NOSAY, 13 disagree strongly with VOTING. It is also seen that of the 271 cases who agree strongly with VOTING, 30 disagree strongly with NOSAY. Are these numbers consistent with an underlying bivariate normal distribution? The answer is given in the next table in the output.

```
                     Correlations and Test Statistics

             (PE=Pearson Product Moment, PC=Polychoric, PS=Polyserial)
                                        Test of Model        Test of Close Fit
  Variable vs. Variable Correlation  Chi-Squ.  D.F. P-Value    RMSEA  P-Value
  -------- --- -------- -----------  --------  ---- -------    -----  -------
    VOTING vs.    NOSAY  0.331 (PC)  216.041    8    0.000     0.129   0.001
  W_A_R_N_I_N_G: Underlying bivariate normality may not hold, see BTS-file

   COMPLEX vs.    NOSAY  0.337 (PC)   76.385    8    0.000     0.074   0.997
   COMPLEX vs.   VOTING  0.284 (PC)   41.882    8    0.000     0.052   1.000
    NOCARE vs.    NOSAY  0.557 (PC)   79.008    8    0.000     0.076   0.996
    NOCARE vs.   VOTING  0.275 (PC)   77.722    8    0.000     0.075   0.996
    NOCARE vs.  COMPLEX  0.459 (PC)   63.037    8    0.000     0.067   1.000
     TOUCH vs.    NOSAY  0.398 (PC)   84.954    8    0.000     0.079   0.989
     TOUCH vs.   VOTING  0.246 (PC)   59.739    8    0.000     0.065   1.000
     TOUCH vs.  COMPLEX  0.357 (PC)   88.841    8    0.000     0.081   0.981
     TOUCH vs.   NOCARE  0.646 (PC)   89.942    8    0.000     0.081   0.978
  INTEREST vs.    NOSAY  0.456 (PC)   79.526    8    0.000     0.076   0.995
  INTEREST vs.   VOTING  0.240 (PC)   76.987    8    0.000     0.074   0.997
  INTEREST vs.  COMPLEX  0.380 (PC)   63.819    8    0.000     0.067   1.000
  INTEREST vs.   NOCARE  0.683 (PC)  116.841    8    0.000     0.094   0.745
  INTEREST vs.    TOUCH  0.686 (PC)   90.641    8    0.000     0.082   0.976
```

For each pair of variables this gives the estimated polychoric correlation, the LR chi-square, the degrees of freedom, and the p-value. It is seen that the hypothesis of underlying bivariate normality is rejected for all pairs of variables.

Models do not have to be true to be useful. It is sufficient that the model holds approximately. If the model holds approximately, it will be rejected by the LR chi-square in large samples. We are not interested in this test as such. We are only interested in obtaining a suitable correlation for ordinal variables. The assumption of underlying bivariate normality is needed to calculate the polychoric correlation. How robust is the polychoric correlation to violations of underlying bivariate normality? This was studied in the dissertation of Ana Maria Quiroga (1992) who found that the polychoric correlation is very robust to violations of underlying bivariate normality but the LR chi-square is very sensitive. This has led me to consider an alternative test based on the non-central, rather than the central, chi-square distribution.

I have developed an RMSEA measure of population discrepancy similar to Steiger's (1990) RMSEA measure for structural equation models. By further simulation studies based on varying degree of underlying bivariate non-normality I have found that there are no serious effects of non-normality unless RMSEA is larger than 0.1. The last two columns of the table give the value of RMSEA and the P-value for the test of the hypothesis that the population value of RMSEA is less than 0.1. By these criteria it is seen that the hypothesis of approximate underlying bivariate normality is rejected only for the pair NOSAY vs VOTING. I will therefore take a closer look at the data for this pair.

If the P-value for the test of approximate underlying normality is less than 0.05, PRELIS writes some further information to a file **inputfilename.BTS**, where inputfilename is the name of the input file. For example, if the input file for PRELIS Run 4 is **ORD21.PRL**, the BTS file will be **ORD21.BTS**. In this case, the BTS file gives four bivariate tables for NOSAY vs VOTING. These will be explained in turn.

The first table is the contingency table for NOSAY and VOTING.

```
    Observed Frequencies
                    AS        A        D       DS    | Rowsum
    --------------------------------------------------------
      AS  |         69       51       27       13    |    160
       A  |         80      297       85        9    |    471
       D  |         92      275      413       24    |    804
      DS  |         30       12       51       26    |    119
    --------------------------------------------------------
    Colsum |        271      635      576       72    |   1554
```

The second table gives the expected frequencies under underlying bivariate normality. Note that the univariate margins fit exactly. Note also that the observed frequencies overestimate the expected frequencies in the (AS, DS) and (DS, AS) cells.

```
    Expected Frequencies
                    AS        A        D       DS    | Rowsum
    --------------------------------------------------------
      AS  |       56.4     70.8     31.4      1.4    |  160.0
       A  |      107.1    212.1    141.4     10.4    |  471.0
       D  |      101.5    318.5    339.6     44.4    |  804.0
      DS  |        6.0     33.6     63.6     15.8    |  119.0
    --------------------------------------------------------
    Colsum |      271.0    635.0    576.0     72.0    | 1554.0
```

The next table gives the cell contributions to the LR chi-square statistic. The LR cell contribution in row i and column j is

$$2Np_{ij} \ln(p_{ij} / \hat{\pi}_{ij}).$$

The sum of these cell contributions is the LR chi-square statistic (7) previously given in the output as 222.381. Note that the LR-contributions can be positive or negative. A positive value indicates that the observed value overestimates the corresponding expected value, whereas a negative value indicates that the observed value underestimates the corresponding expected value.

```
LR Contributions
                 AS        A        D       DS    | Rowsum
----------------------------------------------------------
    AS  |      27.9    -33.5     -8.1     57.6    |    43.8
     A  |     -46.6    199.9    -86.6     -2.6    |    64.2
     D  |     -18.1    -80.8    161.6    -29.5    |    33.2
    DS  |      96.1    -24.7    -22.5     25.8    |    74.8
----------------------------------------------------------
 Colsum |      59.3     61.0     44.4     51.3    |   216.0
```

The next table gives the cell contributions to the GF chi-square statistic (9). The GF cell contribution in row i and column j is

$$N\left(p_{ij} - \hat{\pi}_{ij}\right)^2 / \hat{\pi}_{ij}.$$

These are always non-negative, so they do not provide any information about over- or underestimation. The sum of these cell contributions is the GF-statistic which in this case is 325.4, quite different from the LR-statistic. Here it can be seen more clearly that most of the lack of fit is associated with the (AS,DS) and (DS, AS) cells. It is also seen that the bad fit is associated with the AS and DS categories.

```
GF Contibutions
                 AS        A        D       DS    | Rowsum
----------------------------------------------------------
    AS  |       2.8      5.5      0.6     94.5    |   103.5
     A  |       6.8     34.0     22.5      0.2    |    63.5
     D  |       0.9      5.9     15.9      9.4    |    32.0
    DS  |      94.9     13.9      2.5      6.5    |   117.8
----------------------------------------------------------
 Colsum |     105.5     59.3     41.5    110.6    |   316.8
```

The best way to locate the source of bad fit is to examine the standardized residuals. A residual for a cell can be defined as the square root of a GF contribution, i.e.,

$$\frac{n_{ij} - N\hat{\pi}_{ij}}{\sqrt{N\hat{\pi}_{ij}}} = \sqrt{N}\frac{p_{ij} - \hat{\pi}_{ij}}{\sqrt{\hat{\pi}_{ij}}}. \tag{10}$$

It turns out that the asymptotic variance of (10) is not 1 but smaller than 1. The asymptotic variance of (10) is (see e.g., Rao (1965), eq. (6b.3.2))

$$v_{ij} = 1 - \pi_{ij} - \sum_s \sum_t \left(1/\pi_{ij}\right)\left(\partial \pi_{ij}/\partial \theta_s\right)\left(\partial \pi_{ij}/\partial \theta_t\right)e^{st},$$

where e^{st} is an element of the inverse of the information matrix

$$\mathbf{E} = \sum_{ij}\left(1/\pi_{ij}\right)\left(\partial \pi_{ij}/\partial \mathbf{\theta}\right)\left(\partial \pi_{ij}/\partial \mathbf{\theta}'\right).$$

The residual in (10) divided by the square root of v_{ij} gives the standardized residual. The standardized residuals are given in the last table of the BTS file. It is seen that several standardized residuals are too large or too small. The two largest standardized residuals occur for the (AS,DS) and (DS, AS) cells.

```
Standardized Residuals
              AS        A        D        DS      | Rowsum
------------------------------------------------------------
    AS  |    2.6     -3.4     -1.0     10.2      |     8.4
     A  |   -3.9      9.6     -7.3     -0.5      |    -2.1
     D  |   -1.5     -4.6      8.5     -5.1      |    -2.8
    DS  |   10.9     -4.9     -2.5      3.2      |     6.7
------------------------------------------------------------
  Colsum |   8.1     -3.3     -2.4      7.7      |    10.2
```

All the tables in the BTS file suggest that there are too many cases in the (AS, DS) and (DS, AS) cells. This is in line with the previous argument that VOTING is an ambiguous item.

What should one do when underlying bivariate normality does not even hold approximately? From a practical point of view there are three things that can be done.

- Reduce the number of categories. In this case one could collapse the two categories *agree strongly* and *agree* into one category and the two categories *disagree strongly* and *disagree* into one category. This can be done for any of the variables. If it is done for all variables, all variables will be dichotomous. The model of underlying bivariate normality then becomes saturated as the total number of parameters (3) equals the number of independent cells in the contingency table.
- Eliminate the most offending variables thereby obtaining more homogeneity for the retained variables. In this case I would recommend deleting the variable VOTING since this has an ambiguous question wording.
- If the probability of the various response patterns depends on covariates such as gender, age, income, and education, one can replace the assumption of underlying bivariate normality with the assumption of underlying bivariate normality conditional on the covariates. This is a much more flexible assumption. I will consider this case in Section 5.

It is better to be theory-driven than data-driven. So, I choose the second alternative. The VOTING variable can be eliminated in PRELIS or in LISREL. I choose to eliminate it in LISREL. To eliminate it in PRELIS, just add the line

```
    Sdelete VOTING
```

in PRELIS Run 4. Alternatively, one can select the variables to be included with

```
    Select NOSAY COMPLEX - INTEREST
```

At the end of the output file for PRELIS Run 4, the matrix of polychoric correlations is given as

```
Correlation Matrix

             NOSAY    VOTING   COMPLEX    NOCARE     TOUCH  INTEREST
            --------  --------  --------  --------  --------  --------
    NOSAY    1.000
   VOTING    0.326     1.000
  COMPLEX    0.242     0.289     1.000
   NOCARE    0.441     0.268     0.440     1.000
    TOUCH    0.337     0.167     0.315     0.663     1.000
 INTEREST    0.399     0.223     0.368     0.711     0.630     1.000
```

Note that the correlations for VOTING are all smaller than all the other correlations. This is another indication that VOTING does not belong with the other items.

6 Fitting and Testing the Model

Do the five efficacy items (after exclusion of VOTING) measure a unidimensional trait? This can be tested by running the following SIMPLIS syntax file (file **ORD22.SPL**):

```
    EFFICACY: LISREL Run 1
    Testing Measurement Model
    Observed Variables: NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
    Correlation Matrix from File EFFICACY.PM
    Asymptotic Covariance Matrix from File EFFICACY.ACP
    Sample Size: 1589
    Latent Variable: Efficacy
    Relationships:
       NOSAY COMPLEX - INTEREST = Efficacy
    Method of Estimation: Weighted Least Squares
    Path Diagram
    End of Problem
```

Note that the selection of variables is automatic, i.e., no selection line is needed. VOTING is included in the correlation matrix and in the asymptotic covariance matrix but since it is not included in the model, it will not be used.

Since PRELIS Run 4 generates a data system file or DSF file for short, see Jöreskog, et al. (2001), p. 169. By reading the data from this file, here assumed to have the name **ORD21.DSF**, one can use the following shorter SIMPLIS command file (file **ORD22A.SPL**):

```
    EFFICACY: LISREL Run 1
    System file from File ORD21.DSF
    Latent Variable: Efficacy
```

```
Relationships:
    NOSAY COMPLEX - INTEREST = Efficacy
Method of Estimation: Weighted Least Squares
Path Diagram
End of Problem
```

Some selected lines from the fit statistics are

<div align="center">Goodness-of-Fit Statistics</div>

```
Degrees of Freedom for C(1),C(6)                    5
Weighted Least Squares Chi-Square (C1)              28.436 (P = 0.0000)
Yuan-Bentler (1997) Chi-Square for C1 (C6)          27.925 (P = 0.0000)
Root Mean Square Error of Approximation (RMSEA)     0.0549
90 Percent Confidence Interval for RMSEA            (0.0364 ; 0.0754)
P-Value for Test of Close Fit (RMSEA < 0.05)        0.304
```

In terms of exact fit, the model is rejected. However, the last three lines indicate that the model might hold approximately in the population, see Browne & Cudeck (1993). But since I want a model that has a better chance of being invariant over time and across countries, I will consider an alternative model. It has been suggested in the political science literature that there are two components of Political Efficacy: Internal Efficacy (here called *Efficacy*) indicating individuals self- perceptions that they are capable of understanding politics and competent enough to participate in political acts such as voting, and External Efficacy (here called *Responsiveness* and abbreviated *Respons*) indicating the belief that the public cannot influence political outcomes because government leaders and institutions are unresponsive (Miller, et al., 1980; Craig & Maggiotto, 1982). With this view, NOSAY and COMPLEX are indicators of *Efficacy* and TOUCH and INTEREST are indicators of *Respons*. The statement NOCARE contains two referents: public officials and people like me. This statement might elicit perceptions of the responsiveness of government officials to public opinion generally, in which case the emphasis is on the political actors, or it might express the opinions of people like me in which case the emphasis is on the respondent. In the first case, NOCARE measures *Respons*; in the second case, it measures *Efficacy*. I will therefore consider NOCARE as a complex variable, i.e., as a variable measuring both *Efficacy* and *Respons* or a mixture of them. To estimate this model, use the following SIMPLIS input file (file **ORD23A.SPL**) or, alternatively, **ORD23.SPL**:

```
EFFICACY: LISREL Run 2
Testing Measurement Model 2
Observed Variables: NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
Correlation Matrix from File EFFICACY.PM
Asymptotic Covariance Matrix from File EFFICACY.ACP
Sample Size: 1589
Latent Variable: Efficacy Respons
Relationships:
    NOSAY COMPLEX NOCARE = Efficacy
    NOCARE TOUCH INTEREST = Respons
Method of Estimation: Weighted Least Squares
Path Diagram
End of Problem
```

This gives the following test statistics

```
                        Goodness-of-Fit Statistics

Degrees of Freedom for C(1),C(6)                        3
Weighted Least Squares Chi-Square (C1)                  1.832 (P = 0.6079)
Yuan-Bentler (1997) Chi-Square for C1 (C6)             1.830 (P = 0.6083)

Root Mean Square Error of Approximation (RMSEA)         0.0
90 Percent Confidence Interval for RMSEA               (0.0 ; 0.0350)
P-Value for Test of Close Fit (RMSEA < 0.05)           0.995
```

indicating that the model fits very well. The path diagram with parameter estimates is shown in Fig. 3. The path from *Respons* to NOCARE is not significant. This does not mean that this path should be eliminated if one believes it exists.
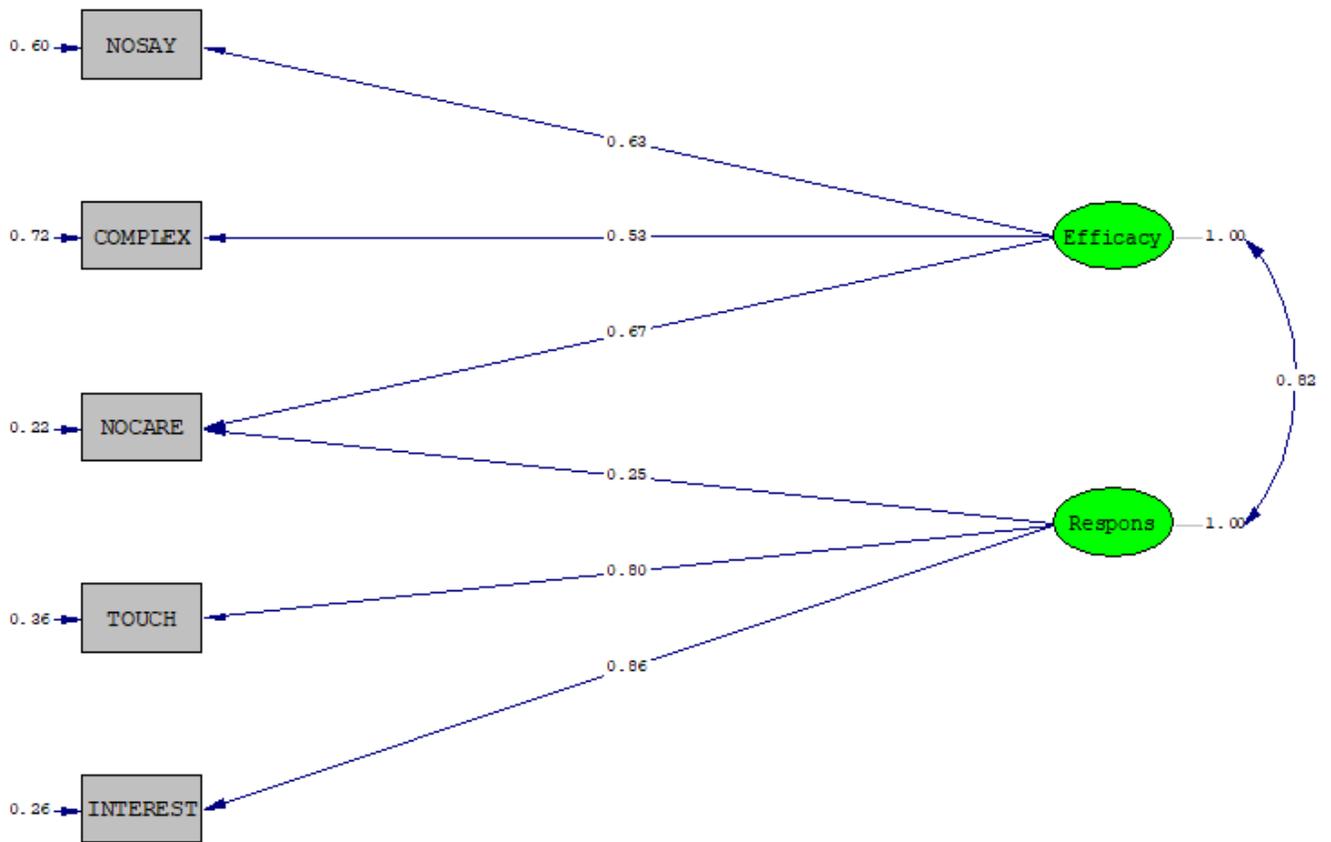
Chi−Square=1.83, df=3, P−value=0.60789, RMSEA=0.000

Figure 3: Path Diagram with Parameter Estimates

I will investigate this model and its invariance over time in the next examples in this set, and its invariance across countries in the fourth example.

7   How to Become a Chi-square Collector

It is easy to become a chi-square collector. Each output from LISREL gives you at least two χ^2 s, sometimes four or five. I have many "wrong" ones and some "right" ones in my collection. To get some of each kind, add the line

```
Method: Maximum Likelihood
```

before the End of Problem line. This means that the ML method will be used to fit the model but standard errors and chi-squares will be corrected for non-normality using the asymptotic covariance matrix provided. This has the advantage that the asymptotic covariance matrix need not be inverted. This method works better with smaller sample sizes. Simulation studies suggest that this may work with samples as small as 200.
The output gives four different chi-squares as

```
Degrees  of  Freedom  = 3
Minimum  Fit  Function  Chi-Square  =  3.85  (P  =  0.28)
Normal Theory Weighted Least Squares Chi-Square = 3.83 (P = 0.28)
Satorra-Bentler Scaled Chi-Square = 1.44 (P = 0.70)
Chi-Square Corrected for Non-Normality = 1.38 (P = 0.71)
```

The first two are the "wrong" ones and the last two are the "right" ones[1]. The first two (3.85 and 3.83) assume normality but normality does not hold here. The last two chi-squares (1.44 and 1.38) take non-normality into account by using asymptotic covariance matrix. The third chi-square (1.44) does not require this matrix to be inverted whereas the fourth one (1.38) does. It is the third one that seems to work best in small samples. For further explanation see Jöreskog, et al. (2001), Chapter 4 and Appendix A.

8   Alternative Parameterization

A weakness of the approach outlined in Sections 1 and 2 is that all underlying variables are standardized to zero means and unit standard deviations. But when the response alternatives are the same for several variables, differences in distributions of these variables may reflect differences in means and/or variances of the underlying variables.

---

[1] You might say that it does not matter if you choose the wrong chi-square since it does not change the conclusion about the fit of the model. However, the "wrong" ones are in fact about 80% larger than the "right" ones. In many cases these differences are large enough to make a real difference in conclusion. I suggest that you do the same analysis with LISREL Run 1. Then the "right" ones suggest that fit is almost OK but the "wrong" ones suggest that the model does not fit at all.

To use such differences, another parameterization of the underlying variables is necessary. This is obtained by putting AP on the OU line in older versions of PRELIS. This parameterization will now automatically be used if MA = CM (or MA = MM or MA = AM). We refer to the default parameterization as the *Standard Parameterization* and the AP parameterization as the *Alternative Parameterization*.

The two parameterizations can be explained as follows. The variable z^* underlying the ordinal variable z is determined only up to a monotonic transformation. If we want to retain normality of the underlying variable, the transformation must be linear. In principle, one can make an arbitrary linear transformation of the underlying variable. If the number of categories $m > 3$, one such transformation is obtained by specifying that $\tau_1 = 0$ and $\tau_2 = 1$. Then the mean and standard deviation σ of z^* can be defined instead.

The parameters of the two parameterizations are given in the following table.

Parameterization	Mean	Std. Dev.	Thresholds				
Standard	0	1	τ_1	τ_2	τ_3	$\cdots$	τ_{m-1}
Alternative	μ^*	σ^*	0	1	τ_3^*	$\cdots$	τ_{m-1}^*

where

$$\mu^* = -\tau_1 / (\tau_2 - \tau_1), \qquad \sigma^* = 1/(\tau_2 - \tau_1),$$
$$\tau_i^* = (\tau_i - \tau_1)/(\tau_2 - \tau_1), \qquad i = 3, 4, ..., m-1.$$

It should be emphasized that the two parameterizations are equivalent in the sense that there is a one-to-one correspondence between the two sets of parameters. This correspondence carries further as will be illustrated with the Efficacy example.

For the Alternative Parameterization to be meaningful there must be at least three categories. If there are only two categories there is only one threshold and it is impossible to estimate both the mean and the standard deviation of the underlying variable. In this case, PRELIS will fix the threshold at 0, the standard deviation at 1 and estimate the mean.

Before I return to the Efficacy example, I will illustrate the Alternative Parameterization using a small data set consisting of two ordinal variables on a 5-point scale[2]. The contingency table and the marginal distributions are given in the following table.

---

[2] Rolf Steyer generated 5000 observations on these two variables, but for convenience I have converted the data to a 5 x 5 contingency table.

	X2					
X1	1	2	3	4	5	
1	95	307	315	291	56	1054
2	61	353	520	687	236	1857
3	22	172	248	478	227	1147
4	8	53	138	393	232	824
5	0	6	15	44	53	118
	176	891	1236	1893	804	5000

If these two variables are on the same response scale, there is a shift in the distribution: For X1most people are in the low end of the scale whereas for X2most people are in the high end of the scale. How can this fact be represented?

To read the data in PRELIS, one can use a frequency data file **TWOVARS.FRQ** consisting of 24 lines as follows:

```
  85 1 1
 307 1 2
 315 1 3
 291 1 4
  56 1 5
  61 2 1
 353 2 2
 520 2 3
 687 2 4
 236 2 5
  22 3 1
 172 3 2
 248 3 3
 478 3 4
 227 3 5
   8 4 1
  53 4 2
 138 4 3
 393 4 4
 232 4 5
   6 5 2
  15 5 3
  44 5 4
  53 5 5
```

Reading this is much faster than reading the data file with 5000 lines. Now run the following PRELIS syntax file (file **TWOVARS.PRL**):

```
DA NI=3
RA=TWOVARS.FRQ
LA
FREQ X1 X2
WEIGHT FREQ
OU MA=CM AP
```

The output reveals the following

```
   Univariate Marginal Parameters

   Variable     Mean St. Dev.   Thresholds
   --------     ---- --------   ----------
        X1    0.795    0.989    0.000    1.000    1.669    2.757
        X2    1.783    0.986    0.000    1.000    1.686    2.760
```

from which it is seen that there is a shift in the mean of the underlying variables of about one unit, whereas the standard deviations and the thresholds are roughly the same.

I now return to the Efficacy example. Run the following PRELIS input (file **ORD24.PRL**):

```
EFFICACY: PRELIS Run 5
Computing Polychoric Correlations and Asymptotic Covariance Matrix
Data Ninputvariables = 6
Labels
NOSAY VOTING COMPLEX NOCARE TOUCH INTEREST
RA=EFFICACY.IMP
CLabels NOSAY - INTEREST 1=AS 2=A 3=D 4=DS
Output MA=CM CM=EFFICACY.CM AC=EFFICACY.ACC
```

This is the same as PRELIS Run 4 except for the last line. Here we are computing the covariance matrix and its asymptotic covariance matrix instead of the polychoric correlation matrix and its asymptotic covariance matrix. The output gives the covariance matrix, the mean vector and the standard deviations as

```
Covariance Matrix

              NOSAY      VOTING     COMPLEX     NOCARE      TOUCH    INTEREST
            --------    --------    --------   --------   --------   --------
   NOSAY     0.948
   VOTING    0.281      0.761
  COMPLEX    0.209      0.158       0.405
   NOCARE    0.460      0.204       0.248      0.720
    TOUCH    0.258      0.143       0.152      0.366      0.445
 INTEREST    0.341      0.161       0.186      0.445      0.351      0.591


 Total Variance = 3.870 Generalized Variance = 0.00619


 Largest Eigenvalue = 2.074 Smallest Eigenvalue = 0.158


 Condition Number = 3.628


 Means

              NOSAY      VOTING     COMPLEX     NOCARE      TOUCH    INTEREST
            --------    --------    --------   --------   --------   --------
              1.231      0.817       0.538      0.868      0.649      0.752
```

Standard Deviations

	NOSAY	VOTING	COMPLEX	NOCARE	TOUCH	INTEREST
	--------	--------	--------	--------	--------	--------
	0.974	0.872	0.637	0.849	0.667	0.769

The covariance matrix obtained in PRELIS Run 5 is just a scaling of the matrix of polychoric correlations obtained in PRELIS Run 4 using the standard deviations from Run 5 as scale factors. These scale factors are also used in the computation of the asymptotic covariance matrix. The mean vector and covariance matrix can be used in LISREL in the same way as for continuous variables. The alternative parameterization is most useful in combination with equal thresholds or fixed thresholds. This will be the topics in the next two examples.

For the moment it is sufficient to do the following. Replace the two lines in **ORD23.SPL**, see file **ORD25.SPL**:

```
Correlation  Matrix  from  File  EFFICACY.PM   Asymptotic Covariance Matrix from File
EFFICACY.ACP
```

with

```
Covariance Matrix from File EFFICACY.CM
Asymptotic Covariance Matrix from File EFFICACY.ACC
```

Alternatively, you can replace **ORD21.DSF** in **ORD23A.SPL** with **ORD24.DSF**, see **ORD25A.SPL**. Verify that the factor loadings and error variances are different but the standardized solution and all fit statistics are the same.