# Announcing LISREL 11

## Contents

# 1. Introduction

LISREL 11 introduces several new features that were not available in previous versions.

Variable names can now be up to 16 characters long compared to 8 characters in previous versions, offering considerable flexibility in the naming of variables. The length of the path names that LISREL can accommodate has also been extended to 192 characters. The path diagrams can also accommodate the longer variable names in the display. Naming conventions and examples of the use of longer path names are given in Section 2.

The *.PTH or path diagram file is now self contained, allowing users to share these files with fellow researchers. It offers a cleaner display and users will no longer be prompted to save this file if no changes have been made to the path diagram. Path diagrams for adaptive quadrature analyses now include the display of the – 2 *ln* L and number of parameters estimated (nfree) on the path diagram, allowing comparison of nested models through the calculation of a chi-square statistic to evaluate improvement in model fit over the models.

To avoid accidentally running the wrong program, only the **Run LISREL** button will be enabled for files with file extension ***.lis** (LISREL syntax), ***.spl** (SIMPLIS syntax), ***.lpj** (LISREL syntax generated through the GUI), and ***.spj** (SIMPLIS syntax generated through the GUI). The **Run PRELIS** button will become active when a ***.prl** file (PRELIS, Multilevel, Multilevel GLIM, Survey GLIM syntax files) is active. If a user used a different file extension for a syntax file, for example ***.inp**, both the **Run LISREL** and **Run PRELIS** options will be disabled and the user would have to rename the syntax file to have the appropriate file extension.

Two-stage multiple imputation SEM for ordinal variables is now available. In previous versions of LISREL, the MCMC multiple imputation method for continuous variables was used to impute missing data values for ordinal variables under the assumption of underlying normal distributions. The new MCMC multiple imputation method for ordinal variables avoids the underlying normality assumption. For more information on the advantages of the two-stage multiple imputation SEM for ordinal variables see Chuang & Cai (2019). This procedure yields more reliable fit statistics for ordinal variables compared to the previous versions. This feature is covered in Section 3.

# 2. Naming convention

LISREL 11 allows users to use variable names up to 16-characters long. In the sections to follow, the rules for variable naming and examples of use are given.

- Variable names are case sensitive.
- When a blank space is used as part of the name, the entire name should be enclosed in single quotes. For example, the name 'Visual Percept' will work, but Visual Percept (without quotes) will not as LISREL will assume the blank space in the name to be the space between two successive variable names. Likewise, 'Visual Perception' will not work as the name is 17 characters long.
- All variables, observed or latent, can have names up to 16-characters long.
- The use of special characters, such as $, * , + etc. are allowed provided the name is enclosed in quotes. A name such as Visual-Percept will not work due to the inclusion of "-". To use this name, it should be given as 'Visual-Percept'.
- When neither blank spaces or special characters are used as part of a variable name, no quotes are needed. For example, VisualPerception can successfully be used as a variable name.
- Labels can carry over lines, with a maximum of 256 characters per line.

## 2.1 Using imported data

If data are imported from an external file and variables have names longer than 16 characters, LISREL will truncate the names to 16 characters. Should the first 16 characters of multiple variables in the imported data be the same, LISREL will stop with an error message indicating duplication.

## 2.2 Using raw data

If raw data or correlation matrices are used, observed variable names should be given as Observed Variables in SIMPLIS and using the LA command in LISREL. Latent variable names can also be read from an external file in the same way as in previous versions. The best way to read names from an external file is to leave a space between variable names.

## 2.3 Examples

Note that the examples to follow can be found on the LISREL Examples page.

### 2.3.1 SIMPLIS example: multiple group analysis

An interesting example of multi-sample analyses is given by Mare & Mason (1981) in a study of the reliabilities of son's and parents' reports on father's and mother's education and occupation. They report the covariance matrices given in Table 1 below for six variables and three populations. The sample size is 80 for each group.

#### Table 1: Covariance matrices for SAT Verbal and Math sections

Covariance matrix for

| Sixth grade | | | | | | |
|---|---|---|---|---|---|---|
| Tests | 'Sons father educ' | 'Sons mother educ' | 'Son Father Occup' | 'Father Own Educ' | 'Mother Own Educ' | 'Father Own Occup' |
| 'Sons father educ' | 5.86 | | | | | |
| 'Sons mother educ' | 3.12 | 3.32 | | | | |
| 'Son Father Occup' | 35.28 | 23.85 | 622.09 | | | |
| 'Father Own Educ' | 4.02 | 2.14 | 29.42 | 5.33 | | |
| 'Mother Own Educ' | 2.99 | 2.55 | 19.20 | 3.17 | 4.64 | |
| 'Father Own Occup' | 35.30 | 26.91 | 465.62 | 31.22 | 23.38 | 546.01 |

| Ninth grade | | | | | | |
|---|---|---|---|---|---|---|
| Tests | 'Sons father educ' | ‘Sons mother educ’ | 'Son Father Occup' | 'Father Own Educ' | 'Mother Own Educ' | 'Father Own Occup' |
| 'Sons father educ' | 8.20 | | | | | |
| ‘Sons mother educ’ | 3.47 | 4.36 | | | | |
| 'Son Father Occup' | 45.65 | 22.58 | 611.63 | | | |
| 'Father Own Educ' | 6.39 | 3.16 | 44.62 | 7.32 | | |
| 'Mother Own Educ' | 3.22 | 3.77 | 23.47 | 3.33 | 4.02 | |
| 'Father Own Occup' | 45.58 | 22.01 | 548.00 | 40.99 | 21.43 | 585.14 |

| Twelfth grade | | | | | | |
|---|---|---|---|---|---|---|
| Tests | 'Sons father educ' | ‘Sons mother educ’ | 'Son Father Occup' | 'Father Own Educ' | 'Mother Own Educ' | 'Father Own Occup' |
| 'Sons father educ' | 5.74 | | | | | |
| ‘Sons mother educ’ | 1.35 | 2.49 | | | | |
| 'Son Father Occup' | 39.24 | 12.73 | 535.30 | | | |
| 'Father Own Educ' | 4.94 | 1.65 | 37.36 | 5.39 | | |
| 'Mother Own Educ' | 1.67 | 2.32 | 15.71 | 1.85 | 3.06 | |
| 'Father Own Occup' | 40.11 | 12.94 | 496.86 | 38.09 | 14.91 | 538.76 |

The variables are:

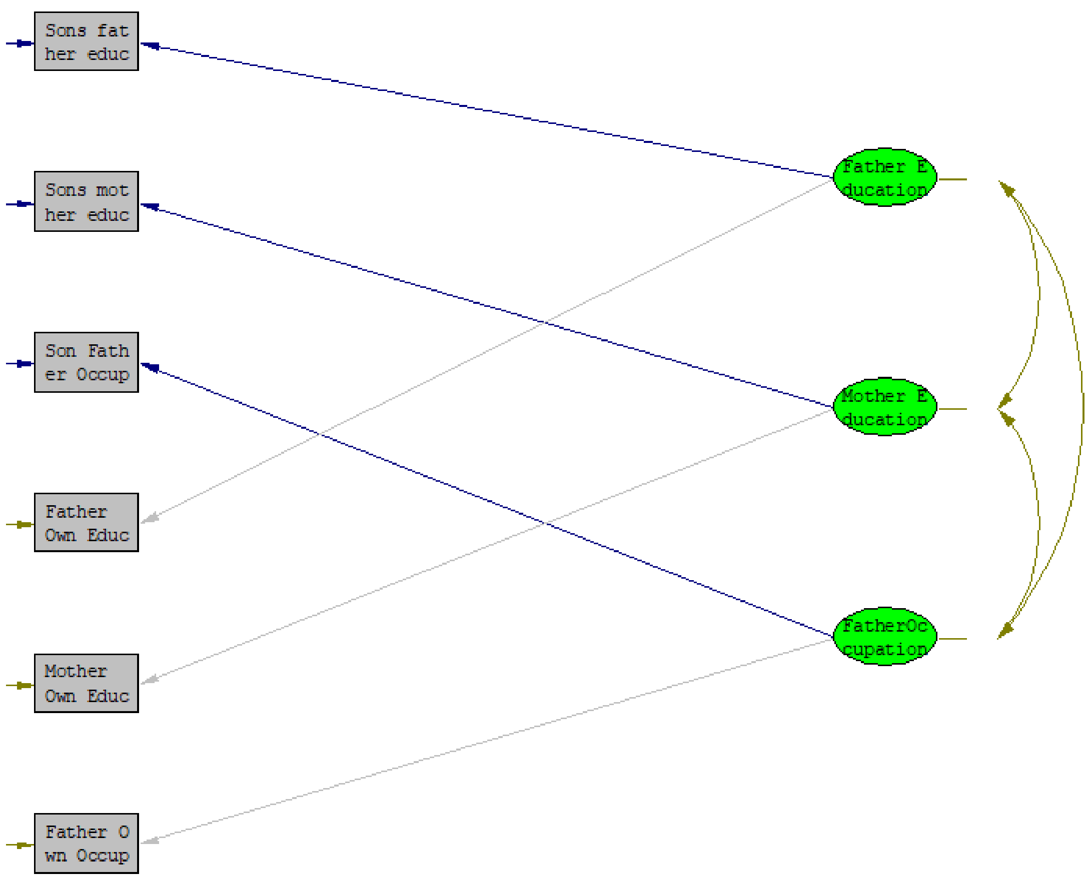
- 'Sons father educ' = Son’s report of father’s education,
- ‘Sons mother educ’ = Son’s report of mother’s education,
- 'Son Father Occup' = Son’s report of father’s occupation,
- 'Father Own Educ' = Father’s report of his own education,
- 'Mother Own Educ' = Mother’s report of her own education, and

- 'Father Own Occup'= Father's report of his own occupation.

The three populations are:

- Group 1: Sixth graders
- Group 2: Ninth graders
- Group 3: twelfth graders

The model is shown in the figure below, where the latent variables 'Father Education', 'Mother Education' and FatherOccupation represent the true father's education, mother's education, and father's occupation, respectively. The use of single quotes with these variable names is necessary as the names have blank spaces in them.




It seems reasonable to assume that the covariance matrix of the latent variables 'Father Education', 'Mother Education' and FatherOccupation are the same in all groups and that the error variances of variables 'Father Education', 'Mother Education' and FatherOccupation should be the same in all groups but that the error variances of 'Sons father educ', 'Sons mother educ' and 'Son Father Occup' should vary over groups. Such a model may be specified as follows (**EX11A_16.SPL**):

```
Group 1: Reports of Parental Socioeconomic Characteristics - Grade 6
Observed Variables: 'Sons father educ' 'Sons mother educ' 'Son Father Occup' 'Father Own Educ'
'Mother Own Educ' 'Father Own Occup'
Covariance Matrix
5.86 3.12 3.32 35.28 23.85 622.09 4.02 2.14 29.42 5.33
2.99 2.55 19.20 3.17 4.64 35.30 26.91 465.62 31.22 23.38 546.01
Sample Size: 80
```

```
Latent Variables: 'Father Education' 'Mother Education' FatherOccupation
'Sons father educ' = 'Father Education'
'Sons mother educ' = 'Mother Education'
'Son Father Occup' = FatherOccupation
'Father Own Educ' = 1*'Father Education'
'Mother Own Educ' = 1*'Mother Education'
'Father Own Occup' = 1*FatherOccupation

Group 2: Reports of Parental Socioeconomic Characteristics - Grade 9
Covariance Matrix
8.20 3.47 4.36 45.65 22.58 611.63 6.39 3.16 44.62 7.32
3.22 3.77 23.47 3.33 4.02 45.58 22.01 548.00 40.99 21.43 585.14
'Sons father educ' = 'Father Education'
'Sons mother educ' = 'Mother Education'
'Son Father Occup'= FatherOccupation
Let the Error Variances of 'Sons father educ' - 'Son Father Occup' be free

Group 3: Reports of Parental Socioeconomic Characteristics - Grade 12
Covariance Matrix
5.74 1.35 2.49 39.24 12.73 535.30 4.94 1.65 37.36 5.39
1.67 2.32 15.71 1.85 3.06 40.11 12.94 496.86 38.09 14.91 538.76
'Sons father educ' = 'Father Education'
'Sons mother educ' = 'Mother Education'
'Son Father Occup'= FatherOccupation
Let the Error Variances of 'Sons father educ'  - 'Son Father Occup' be free
Path Diagram
End of Problem
```

Note the following:

- Since the sample size is the same for all groups, it need only be specified for the first group.
- The units of measurements of the latent variables 'Father Education', 'Mother Education' and FatherOccupation are defined in the first group by using 'Sons father educ', 'Sons mother educ' and 'Son Father Occup' as reference variables. Since this is not repeated in the second and third group, it stays the same. This defines the units of the latent variables to be the same in all groups, which is essential. Otherwise, it would not make sense to postulate equal variances and covariances for the latent variables.
- Since nothing is specified about the variances and covariances of the latent variables, their covariance matrix will be the same in all groups.

- The inclusion of the line

```
Let the Error Variances of 'Sons father educ' 'Sons mother educ' 'Son Father Occup' be free
```

for the second and third group makes the error variances 'Sons father educ', 'Sons mother educ' and 'Son Father Occup' free in all groups.

The chi-square for this model is

```
Degrees of Freedom for (C1)-(C2)                      36
 Maximum Likelihood Ratio Chi-Square (C1)              78.368 (P = 0.0001)
```

Mare & Mason (1981) considered another model in which the error terms of 'Sons mother educ' and 'Sons father educ' were allowed to correlate in the first two groups but not in the third. To specify this, add the line

```
Set the Error Covariance between 'Sons mother educ' and 'Sons father educ' free
```

in the first two groups, and add the line

```
Set the Error Covariance between 'Sons mother educ' and 'Sons father educ' equal to 0
```

in the third group. The last line is needed, otherwise the error covariance between 'Sons mother educ' and 'Sons father educ' in the third group would be estimated to be equal to that of the second group. The chi-square for this model is

```
Degrees of Freedom for (C1)-(C2)                      34
 Maximum Likelihood Ratio Chi-Square (C1)              52.728 (P = 0.0212)
```

The output file gives the reliabilities (squared multiple correlations) shown in the table below. This is a somewhat remarkable result because in grade 12, the sons' reports of their fathers' education and occupation are more reliable than the fathers' reports of their own education and occupation. In the earlier ages, however, the sons' reports are less reliable that their parents' reports, as expected.

**Table 2: Estimated reliabilities of son's and parents' reports of parental socioeconomic characteristics**

| | 'Sons father educ' | 'Sons mother educ' | 'Son Father Occup' | 'Father Education' | 'Mother Education' | FatherOccupation |
|---|---|---|---|---|---|---|
| Grade 6 | 0.62 | 0.38 | 0.71 | 0.87 | 0.93 | 0.90 |
| Grade 9 | 0.76 | 0.86 | 0.92 | 0.87 | 0.93 | 0.90 |
| Grade 12 | 0.91 | 0.81 | 0.95 | 0.87 | 0.93 | 0.90 |

### 2.3.2 PRELIS example: Polychoric correlation matrix with ordinal variables (MA = PM)

Swedish school children in grade 9 were asked questions about their attitudes on social issues in family, school, and society. Among the questions asked were the following eight items (Hasselrot & Lernberg, 1980).

For me, questions about….

- Human rights
- Equal conditions for all people
- Racial problems
- Equal value of all people
- Euthanasia
- Crime and punishment
- Conscientious objectors
- Guilt and bad conscience

are:

__unimportant   __Not important   __important   __very important

In this example we use a subsample of 200 cases. Response to the eight questions were scored 1,2,3, and 4 (4 = very important). Missing values were scored zero. The data matrix consists of 200 rows and 8 columns. It is stored in data file **DATA.EX2**.

The PRELIS syntax file (**EX2_16.PRL**) is:

```
EXAMPLE 2: ATTITUDES OF MORALITY AND EQUALITY
DA NI=8 NO=200 MI=0 TR=PA
LA
'Human Rights' 'Equal Conditions' 'Racial Problems' 'Equal Value' Euthanasia 'Crime&Punishment'
'Consc. Objectors' Guilt
RA FI=DATA.EX2
OR ALL
OU MA=PM
```

**Note**:

While there is no space in the variable name Crime&Punishment, it is placed in quotes because the special character “&” is used. The same principle requires the use of single quotes for the name ‘Consc. Objectors’ as it has both a “.” and a space in the name.

Partial output for this analysis is shown below.

```
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
 ---------------- --- ----------------  -----------  --------  ---- -------       -----  -------
 Equal Conditions vs.     Human Rights  0.418 (PC)   14.708     8    0.065        0.065   0.853
  Racial Problems vs.     Human Rights  0.221 (PC)   10.872     8    0.209        0.043   0.950
  Racial Problems vs. Equal Conditions  0.202 (PC)    4.933     8    0.765        0.000   0.998
      Equal Value vs.     Human Rights  0.373 (PC)   16.735     8    0.033        0.074   0.777
      Equal Value vs. Equal Conditions  0.635 (PC)   13.097     8    0.109        0.057   0.903
      Equal Value vs.  Racial Problems  0.285 (PC)   10.408     8    0.238        0.039   0.958
       Euthanasia vs.     Human Rights  0.442 (PC)   12.341     8    0.137        0.052   0.921
       Euthanasia vs. Equal Conditions  0.706 (PC)    9.872     8    0.274        0.034   0.968
       Euthanasia vs.  Racial Problems  0.317 (PC)    7.348     8    0.500        0.000   0.990
       Euthanasia vs.      Equal Value  0.692 (PC)   26.238     8    0.001        0.107   0.364
 Crime&Punishment vs.     Human Rights  0.215 (PC)   12.460     8    0.132        0.053   0.918
 Crime&Punishment vs. Equal Conditions  0.236 (PC)   17.785     8    0.023        0.078   0.736
 Crime&Punishment vs.  Racial Problems  0.282 (PC)   14.366     8    0.073        0.064   0.862
 Crime&Punishment vs.      Equal Value  0.423 (PC)   13.672     8    0.091        0.060   0.889
 Crime&Punishment vs.       Euthanasia  0.218 (PC)   13.103     8    0.108        0.056   0.905
 Consc. Objectors vs.     Human Rights  0.190 (PC)   11.504     8    0.175        0.047   0.940
 Consc. Objectors vs. Equal Conditions  0.292 (PC)   10.694     8    0.220        0.041   0.955
 Consc. Objectors vs.  Racial Problems  0.312 (PC)    6.467     8    0.595        0.000   0.994
 Consc. Objectors vs.      Equal Value  0.340 (PC)   14.912     8    0.061        0.066   0.850
 Consc. Objectors vs.       Euthanasia  0.224 (PC)    7.263     8    0.509        0.000   0.991
 Consc. Objectors vs. Crime&Punishment  0.312 (PC)    9.113     8    0.333        0.026   0.977
            Guilt vs.     Human Rights  0.094 (PC)   17.361     8    0.027        0.077   0.749
            Guilt vs. Equal Conditions  0.314 (PC)   15.453     8    0.051        0.069   0.827
            Guilt vs.  Racial Problems  0.239 (PC)    8.626     8    0.375        0.020   0.980
            Guilt vs.      Equal Value  0.320 (PC)    7.392     8    0.495        0.000   0.990
            Guilt vs.       Euthanasia  0.340 (PC)    8.500     8    0.386        0.018   0.983
            Guilt vs. Crime&Punishment  0.207 (PC)    7.595     8    0.474        0.000   0.989
            Guilt vs. Consc. Objectors  0.202 (PC)   12.879     8    0.116        0.055   0.909

           Percentage of Tests Exceeding 0.5% Significance Level:  0.0%
           Percentage of Tests Exceeding 1.0% Significance Level:  0.0%
           Percentage of Tests Exceeding 5.0% Significance Level:  0.0%
```

### 2.3.3 LISREL example: Analysis of reader reliability in essay scoring

In an experiment (Votaw, 1948) to establish methods of obtaining reader reliability in essay scoring, 126 examinees were given a three-part English Composition examination. Each part required the examinee to write an essay, and for each examinee, scores were obtained on the following:

1. the original part 1 essay, represented by the variable ‘Original part1’
2. a handwritten copy of the original part 1 essay (‘Written copy’)
3. a carbon copy of the handwritten copy in (2) (‘Carbon copy’), and
4. the original part 2 essay, represented by the variable ‘Original part2’.

Scores were assigned by a group of readers using procedures designed to counterbalance certain experimental conditions. The investigator would like to know whether, on the basis of this sample of size 126, the four scores can be used interchangeably or whether scores on the copies (2) and (3) are less reliable than the originals (1) and (4).

The covariance matrix of the four measurements is given in the command file below. The hypothesis to be tested in this example is that a one-factor congeneric measurement model describes these data well.

The LISREL command file for this analysis is (**EX31A_16.LIS**):

```
Analysis of Reader Reliability in Essay Scoring Votaw's Data
Congeneric model estimated by ML
DA NI=4 NO=126
LA
'ORIGINAL PART1' 'WRITTEN COPY' 'CARBON COPY' 'ORIGINAL PART2'
CM
25.0704
12.4363 28.2021
11.7257  9.2281 22.7390
20.7510 11.9732 12.0692 21.8707
MO NX=4 NK=1 LX=FR PH=ST
LK
'Essay ability'
PD
OU SE ND=2
```

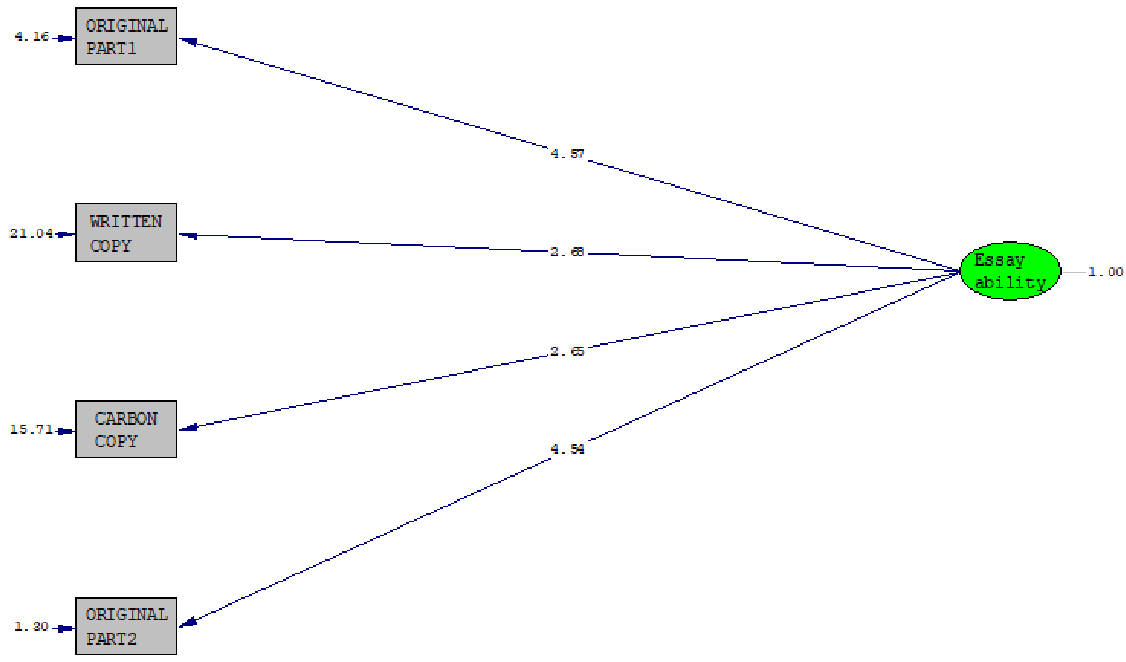
The DA command specifies four observed variables and a sample size of 126; the MA default is assumed, so the covariance matrix will be analyzed. Labels for the input variables follow the LA command. The CM command indicates that a covariance matrix is to be input. Because an external file is not specified, the matrix follows in the command file. A format statement does not appear, so the input is in free format. The MO command specifies four *x*-variables and one latent variable; the elements of $\lambda$ are all free (LX = FR), and the latent variable is standardized (PH = ST). A label for the latent variable follows the LK command. Note that all observed variable names are given in single quotes, as all names have a blank space between parts of the name. The same holds true for the latent variable 'Essay ability'.

In the results of this analysis, the goodness-of-fit statistic

```
 Degrees of Freedom for (C1)-(C2)                     2
 Maximum Likelihood Ratio Chi-Square (C1)             2.298 (P = 0.3169)
 Browne's (1984) ADF Chi-Square (C2_NT)               2.236 (P = 0.3270)
```

indicates that the hypothesis is acceptable.

The path diagram obtained for this model is shown below.




The results under the hypothesis are given in the table below. The three columns of this table can be read off directly from the output for the ML solution. The reliabilities in column 3 appear where the output says "squared multiple correlations for *x*-variables".

**Table: Essay scoring data: results for congeneric model**

| | $\hat{\lambda}_i$ | $se\left(\hat{\lambda}_i\right)$ | $\hat{\rho}_{ii}$ |
|---|---|---|---|
| 1 | 4.57 | 0.36 | 0.83 |
| 2 | 2.68 | 0.45 | 0.25 |
| 3 | 2.65 | 0.40 | 0.31 |
| 4 | 4.54 | 0.33 | 0.94 |

Inspecting the different $\lambda$'s, it is evident that these are different even taking their respective standard errors of estimate into account. Comparing the reliabilities in the last column, one sees that they are high for scores (1) and (4) and low for scores (2) and (3). Thus, it appears that scores obtained from originals are more reliable than scores based on copies.

## 2.3.4 LISREL example: reading variable names from external file

In some cases, it is useful to place the variable names in an external file rather than in the syntax file itself. An example of such an analysis is discussed in this section.

The example is based on data from Duncan, Haller & Portes (1968). Of interest is the way in which a person's peers (e.g., best friends) influence his or her decisions (e.g., choice of occupation). We anticipate that the relation between respondent's ambition (RespAmbition) and best friend's ambition (BFriendAmbition) must be reciprocal. As a test of this model, a sample of Michigan high-school students were paired with their best friends and measured on a number of background variables. In addition, scaled measures of occupational and educational aspiration were obtained to serve as indicators of a latent variable AMBITION.

The observed measures in the study are:

- $x_2$ = respondent's intelligence (RespIntelligence)
- $x_3$ = respondent's socioeconomic status (RespSocEconStat)
- $x_4$ = best friend's socioeconomic status (BFriendSES)
- $x_5$ = best friend's intelligence (BFriendIntel)
- $y_1$ = respondent's occupational aspiration (RespOccupAspire)
- $y_2$ = respondent's educational aspiration (RespEducAspire)
- $y_3$ = best friend's educational aspiration (BFriendEducAsp)
- $y_4$ = best friend's occupational aspiration (BFriendOccupAsp)
- $\eta_1$ = respondent's ambition (RespAmbition)
- $\eta_2$ = best friend's ambition (BFriendAmbition)

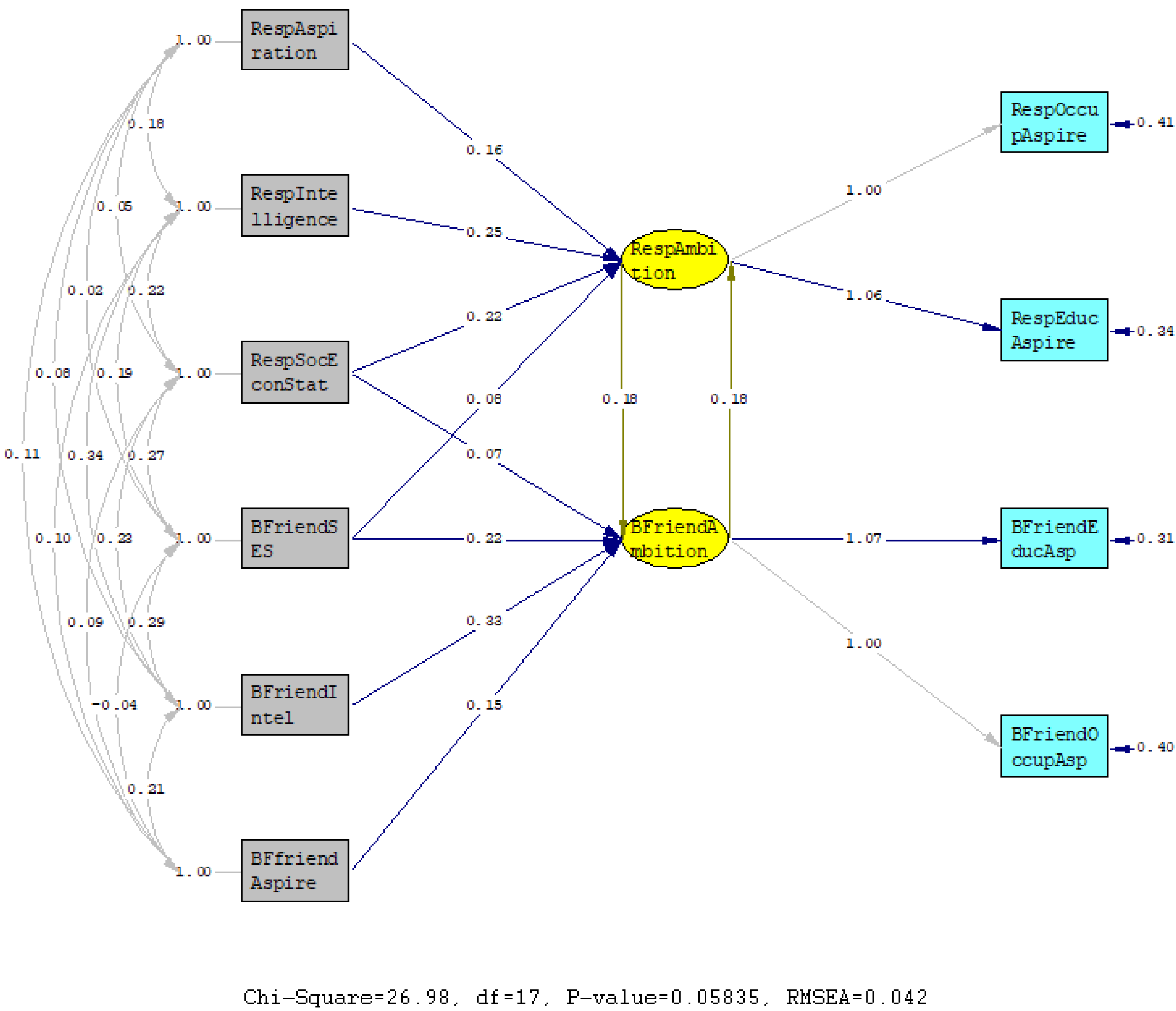
The correlation matrix to be analyzed here is stored in the file **EX55.COR**. Syntax is given in the file **EX55A_16.LIS**:

```
Peer Influences on Ambition: Model with BE(2,1) = BE(1,2) and PS(2,1) = 0
DA NI=10 NO=329
LA FI=EX55_16.LAB
KM FI=EX55.COR
SELECT
4 5 10 9 2 1 3 8 6 7
MO NY=4 NE=2 NX=6 FIXED-X BE=FU PS = DI
LE
RespAmbition BFriendAmbition
FR LY(2,1) LY(3,2) BE(1,2)
FI GA(5)-GA(8)
VA 1 LY(1) LY(8)
EQ BE(1,2) BE(2,1)
OU SE TV EF SS
```

The LAB file, containing the variable names, is as follows:

```
ex55a_16.lab
RespIntelligence RespAspiration  RespSocEconStat RespOccupAspire RespEducAspire
BFriendIntel     BFfriendAspire  BFriendSES BFriendOccupAsp BFriendEducAsp
```

Note that the names are given in free format, separated by a blank between each pair of names. The labels continue in the second line of the file. The path diagram for this model is shown below.



## 2.3.5 Multilevel SEM analysis with structured means

In this example, we consider a multilevel SEM analysis with structured means. The between-schools model is a one factor CFA model with a fixed factor variance, a latent mean, equal intercepts and equal measurement error covariances while the within-schools model is a one factor CFA model with equal measurement error covariances.

The dataset **maths.lsf** is based on a longitudinal study and consists of data from 1721 students nested within 55 schools. This dataset is based on the datasets eg1.sav, eg2.sav and eg3.sav described in Chapter 4 of Raudenbush, S, Bryk, A, Cheong, Y.F., Congdon, R & Du Toit (2011).

The following variables are available:

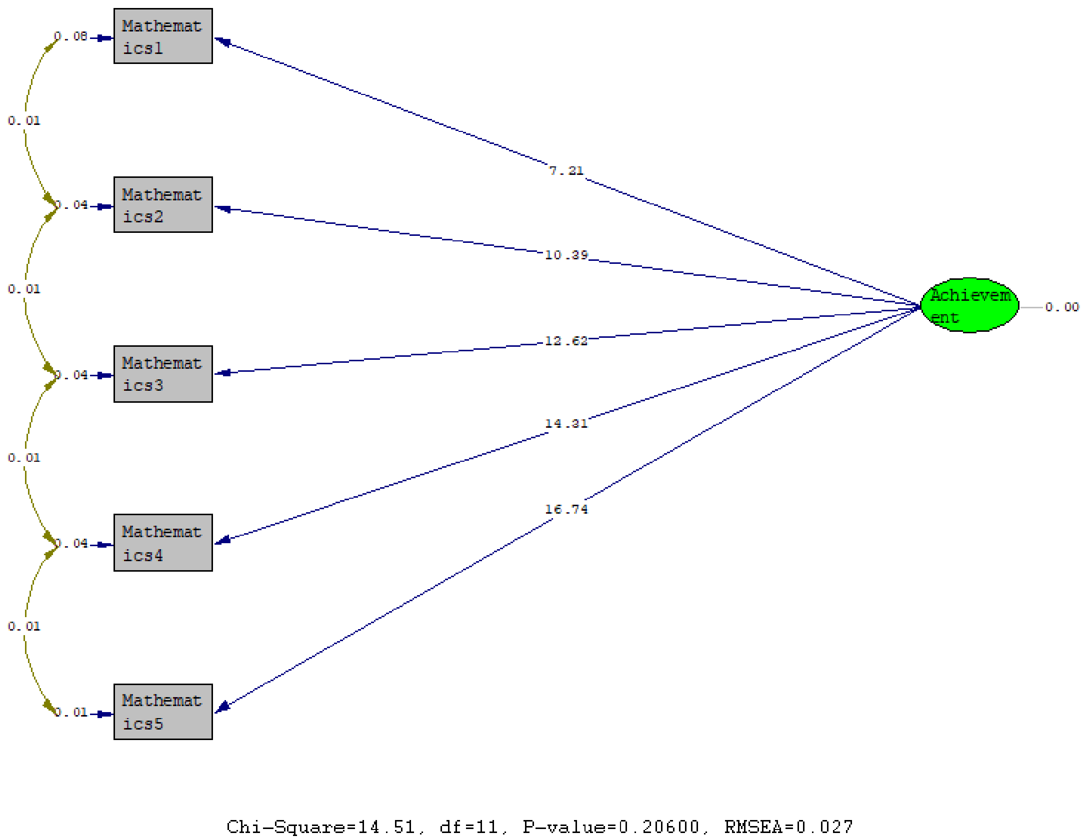
- SchoolID Cluster (level 2) ID
- ChildID Student number
- Retained 1 If retained in the same grade at least once
- ‘Maths score 1’ Score in IRT metric on mathematics test on first measurement occasion
- ‘Maths score 2’ Score in IRT metric on mathematics test on second measurement occasion
- ‘Maths score 3’ Score in IRT metric on mathematics test on third measurement occasion
- ‘Maths score 4’ Score in IRT metric on mathematics test on fourth measurement occasion
- ‘Maths score 5’ Score in IRT metric on mathematics test on fifth measurement occasion
- Gender 1 if female, 0 if male
- ‘Ethnicity 1’ 1 if African American, 0= other
- ‘Ethnicity 2’ 1 if Hispanic, 0= other
- Size School level variable, number of students in school
- ‘Low Income’ School level variable, percent of students from low income families
- Mobility School level variable, percent of students moving during academic year

The model to be fitted is described in the syntax file **math_trend3a_16.lis**.

```
!     STRUCTURED MEANS
! In this example, we specify equal non-random intercepts and a latent mean for the one
! factor CFA model between schools (eq tx(5) tx(4) tx(3) tx(2) tx(1); ka=fr).
TI
DA NI=14 NO=0 NG=2 MA=CM MI=-9
RA FI=maths_16.lsf
$CLUSTER  SchoolID
SE
4 5 6 7 8 /
mo nx=5 nk=1 tx=fr lx=fu,fr td=sy,fi ka=fr ph=sy,fi
lk
Achieve
va 0.001 ph(1,1)
eq tx(5) tx(4) tx(3) tx(2) tx(1)
fr td(1,1) td(2,2) td(3,3) td(4,4) td(5,5)
fr td(2,1) td(3,2) td(4,3) td(5,4)
eq td(2,1) td(3,2) td(4,3) td(5,4)
pd
ou
Group2 : Within Schools
DA NI=14 NO=0 NG=2 MA=CM MI=-9
RA FI=maths_16.LSF
SE
4 5 6 7 8 /
mo nx=5 nk=1 tx=fi lx=fu,fr td=sy,fi ka=fi  ph=sy,fr
lk
```

```
Achieve
va 0.0 ka(1)
va 0.0 tx(5) tx(4) tx(3) tx(2) tx(1)
fi lx(1,1)
va 1.0 lx(1,1)
fr td(1,1) td(2,2) td(3,3) td(4,4) td(5,5)
fr td(2,1) td(3,2) td(4,3) td(5,4)
eq td(2,1) td(3,2) td(4,3) td(5,4)
OU ND=3 MI
```

The path diagram for this model is shown below, followed by selected output.




From the estimated lambdas it appears that the difference in estimates is not linear as these estimates increase monotonically. This would indicate that a linear growth curve model over the period during which measurements were made would probably be more appropriate if we wanted to fit a regression model to these data. The fit statistics given above indicate that the model provides an adequate description of the data.

A section of the output is given below.

Covariance Matrix

| | Mathematics1 | Mathematics2 | Mathematics3 | Mathematics4 | Mathematics5 |
|---|---|---|---|---|---|
| Mathematics1 | 0.138 | | | | |
| Mathematics2 | 0.090 | 0.133 | | | |
| Mathematics3 | 0.127 | 0.161 | 0.251 | | |
| Mathematics4 | 0.123 | 0.158 | 0.250 | 0.290 | |
| Mathematics5 | 0.123 | 0.170 | 0.246 | 0.281 | 0.300 |

Total Variance = 1.113 Generalized Variance = 0.433105D-06

```
 Largest Eigenvalue = 0.970 Smallest Eigenvalue = 0.007

 Condition Number = 11.793

         Means
                Mathematics1   Mathematics2   Mathematics3   Mathematics4   Mathematics5
                ------------   ------------   ------------   ------------   ------------
                      -1.976         -0.929         -0.198          0.356          1.160

 Group2 : Within Schools

         Covariance Matrix
                Mathematics1   Mathematics2   Mathematics3   Mathematics4   Mathematics5
                ------------   ------------   ------------   ------------   ------------
 Mathematics1          0.774
 Mathematics2          0.510          0.875
 Mathematics3          0.610          0.781          1.263
 Mathematics4          0.566          0.701          0.930          1.063
 Mathematics5          0.527          0.668          0.863          0.881          1.035

 Total Variance = 5.010 Generalized Variance = 0.0194

 Largest Eigenvalue = 3.891 Smallest Eigenvalue = 0.160

 Condition Number = 4.930
LISREL Estimates (Maximum Likelihood)

         LAMBDA-X
                 Achievement
                ------------
 Mathematics1          7.205
                     (1.792)
                       4.021

 Mathematics2         10.386
                     (1.670)
                       6.219

 Mathematics3         12.624
                     (1.754)
                       7.197

 Mathematics4         14.306
                     (1.894)
                       7.552

 Mathematics5         16.740
                     (2.190)
                       7.642

Group2 : Within Schools

 Number of Iterations = 54
```

```
 LISREL Estimates (Maximum Likelihood)

         LAMBDA-X
                  Achievement
                 ------------
 Mathematics1           1.000

 Mathematics2           1.233
                      (0.055)
                       22.400

 Mathematics3           1.583
                      (0.071)
                       22.162

 Mathematics4           1.519
                      (0.066)
                       23.044

 Mathematics5           1.458
                      (0.064)
                       22.618

         PHI
                  Achievement
                 ------------
                        0.375
                      (0.035)
                       10.794
```

## 2.3.6 Converting LSF files

This example reads data from an external LSF file. In this case, the names of the variables need to be changed within the LSF file. There are two ways to do so.

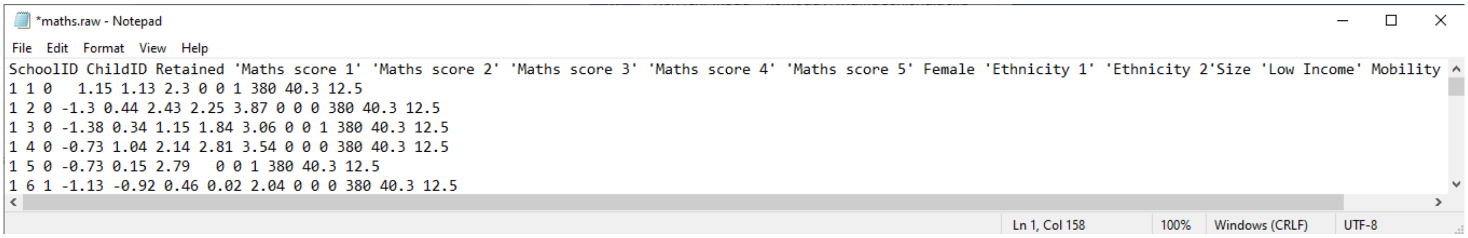
### Using a txt file

This option shows how a txt file can be used to update names in the LSF file. It also illustrates the correct way of adding longer variable names to raw data.

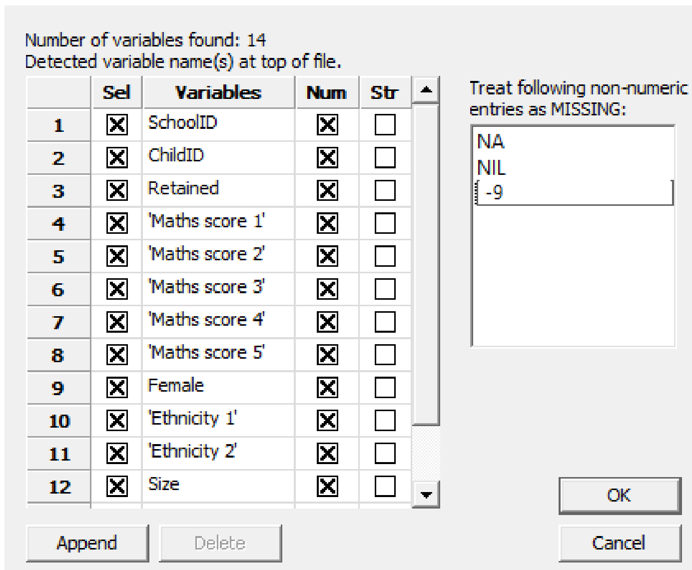
To change the variable names, start by opening the older format **maths.lsf** file used in versions up to LISREL 10.3. Select the **Export** option from the **File** menu and export the contents of **maths.lsf** to the file **maths.raw**. This file is shown below:

```
maths.raw - Notepad
File  Edit  Format  View  Help
SchoolID ChildID Retained Math1 Math2 Math3 Math4 Math5 Female Black Hispanic Size LowIncom Mobility
1 1 0   1.15 1.13 2.3 0 0 1 380 40.3 12.5
1 2 0 -1.3 0.44 2.43 2.25 3.87 0 0 0 380 40.3 12.5
1 3 0 -1.38 0.34 1.15 1.84 3.06 0 0 1 380 40.3 12.5
1 4 0 -0.73 1.04 2.14 2.81 3.54 0 0 0 380 40.3 12.5
1 5 0 -0.73 0.15 2.79   0 0 1 380 40.3 12.5
1 6 1 -1.13 -0.92 0.46 0.02 2.04 0 0 0 380 40.3 12.5
Ln 1, Col 1    100%   Windows (CRLF)   UTF-8
```

Rename the variables as shown below. As most of the names have blanks as part of the names, use single quotes around the variable names as shown.

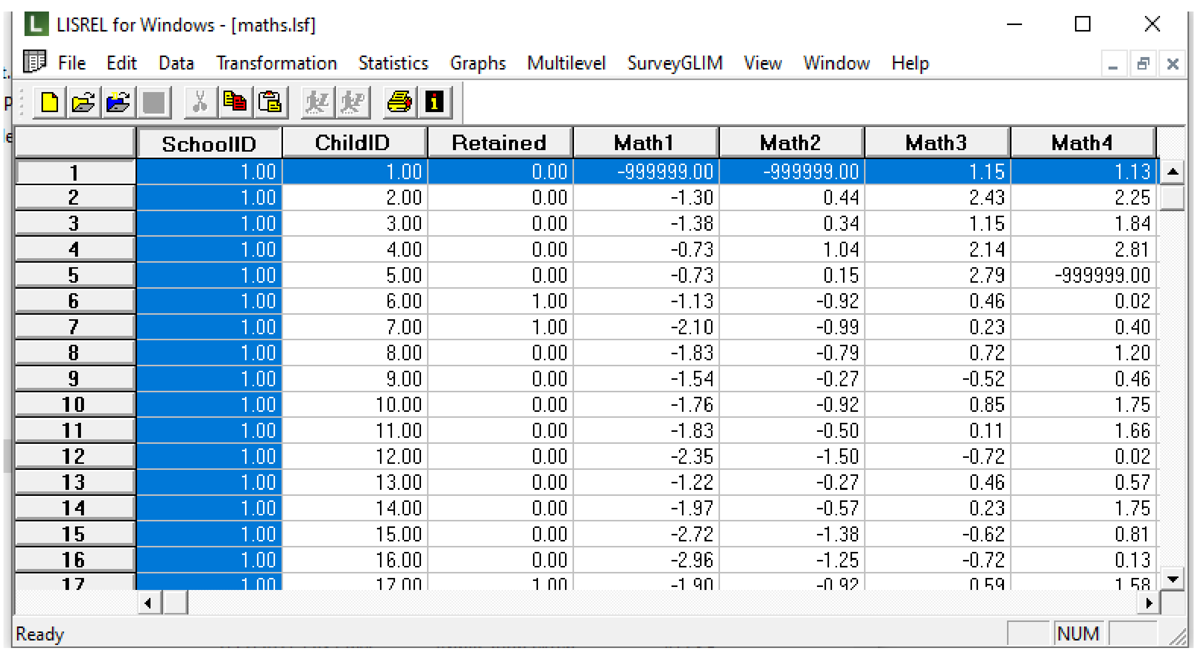



To create an LSF with 16-character variable names, reimport this data into LISREL and save it as **maths_16.lsf**. Remember to address the presence of any missing data when importing. For example, here -9 is defined as a global missing value.



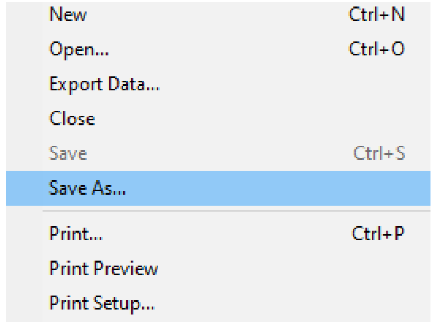

## Using an LSF file

Open the LSF file

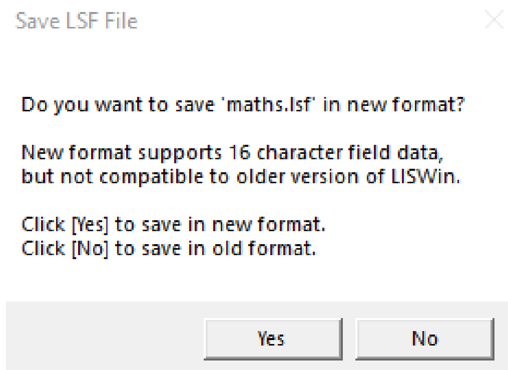



| | SchoolID | ChildID | Retained | Math1 | Math2 | Math3 | Math4 |
|---|---|---|---|---|---|---|---|
| 1 | 1.00 | 1.00 | 0.00 | -999999.00 | -999999.00 | 1.15 | 1.13 |
| 2 | 1.00 | 2.00 | 0.00 | -1.30 | 0.44 | 2.43 | 2.25 |
| 3 | 1.00 | 3.00 | 0.00 | -1.38 | 0.34 | 1.15 | 1.84 |
| 4 | 1.00 | 4.00 | 0.00 | -0.73 | 1.04 | 2.14 | 2.81 |
| 5 | 1.00 | 5.00 | 0.00 | -0.73 | 0.15 | 2.79 | -999999.00 |
| 6 | 1.00 | 6.00 | 1.00 | -1.13 | -0.92 | 0.46 | 0.02 |
| 7 | 1.00 | 7.00 | 1.00 | -2.10 | -0.99 | 0.23 | 0.40 |
| 8 | 1.00 | 8.00 | 0.00 | -1.83 | -0.79 | 0.72 | 1.20 |
| 9 | 1.00 | 9.00 | 0.00 | -1.54 | -0.27 | -0.52 | 0.46 |
| 10 | 1.00 | 10.00 | 0.00 | -1.76 | -0.92 | 0.85 | 1.75 |
| 11 | 1.00 | 11.00 | 0.00 | -1.83 | -0.50 | 0.11 | 1.66 |
| 12 | 1.00 | 12.00 | 0.00 | -2.35 | -1.50 | -0.72 | 0.02 |
| 13 | 1.00 | 13.00 | 0.00 | -1.22 | -0.27 | 0.46 | 0.57 |
| 14 | 1.00 | 14.00 | 0.00 | -1.97 | -0.57 | 0.23 | 1.75 |
| 15 | 1.00 | 15.00 | 0.00 | -2.72 | -1.38 | -0.62 | 0.81 |
| 16 | 1.00 | 16.00 | 0.00 | -2.96 | -1.25 | -0.72 | 0.13 |
| 17 | 1.00 | 17.00 | 1.00 | -1.90 | -0.92 | 0.59 | 1.58 |

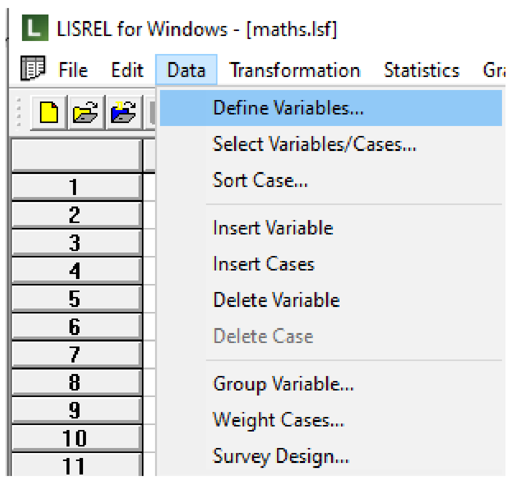
Next, use the **Save As** option on the **File** menu to save the LSF file in the new 16-character format:



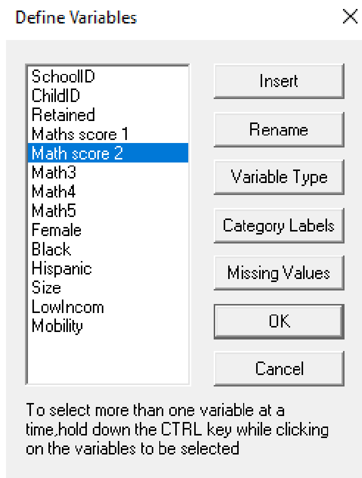

Doing so will lead to the display of a small dialog box on which the user can select the old or new format.




After opting to save it in the new format, use the **Define Variables** option from the **Data** menu to access the **Define Variables** dialog box.




Finally, change the names of the variables on the **Define Variables** dialog box by using the **Rename** option. Click **OK** when done.




### 2.3.7 FIML and missing data example: the assessment of invariance

In practice, many multivariate data sets are observations from several groups. Examples of these groups are genders, languages, political parties, countries, faculties, colleges, schools, etc. For these data sets, it is often of interest to determine whether or not the parameters of the structural equation model for the observed variables are invariant across the groups. The statistical methods for multiple group structural equation modeling may be used to determine whether or not these parameters are invariant across the groups.

LISREL may be used to fit multiple group structural equation models to multiple group data. Traditional statistical methods such as Maximum Likelihood (ML), Robust Maximum Likelihood (RML), Weighted Least Squares (WLS), Diagonally Weighted Least Squares (DWLS), Generalized Least Squares (GLS) and Un-weighted Least Squares (ULS) are available for complete multiple group data while the Full Information Maximum Likelihood (FIML) method is available for incomplete multiple group data. The ML, RML, WLS, DWLS, GLS and ULS methods for multiple group structural equation modeling are described in Jöreskog & Sörbom (1999) while the FIML method is described in Du Toit & Du Toit (2001).

In this example, the FIML estimation method for incomplete data of LISREL is used to fit a measurement model to a multivariate data sets consisting of the simulated scores of a sample of 1250 boys and 1250 girls on six psychological tests. The raw data are given in the LISREL System File **LIS11_MG_BOYS_GIRLS_16**.

Variables of interest are:

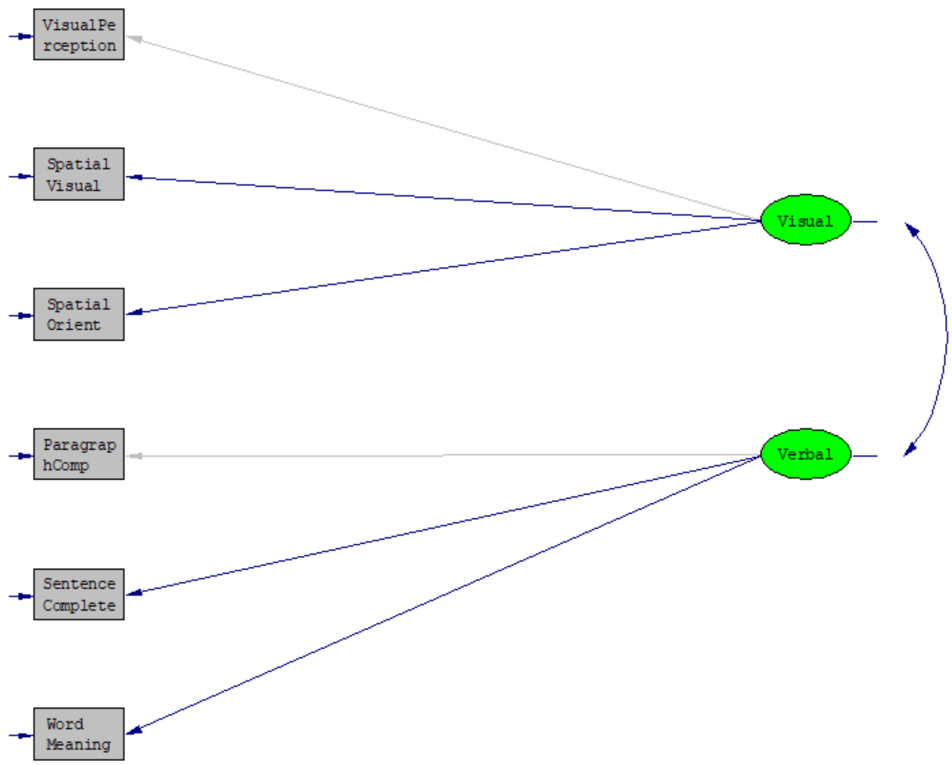
- Visual perception scores (VisualPerception)
- Tests of spatial visualization ('Spatial Visual')
- Test of spatial orientation ('Spatial Orient')
- Paragraph completion score (ParagraphComp)
- Sentence completion score (SentenceComplete)
- Word meaning score ('Word Meaning')

The invariance of a model is often of interest if the sample data consist of data from different groups such as males and females, different political parties, freshmen, sophomores, juniors and seniors, etc. In this section, we illustrate how LISREL can be used to assess various levels of invariance across groups.

Configural invariance is achieved if the model of interest fits across the groups. Although the model is the same across groups, the unknown parameters of the model are assumed to be different across the groups. The multiple group (global) Chi-square test statistic for this multiple group model is used to assess configural invariance. The measurement model in

Figure 1 will now be used to illustrate how the multiple group feature of LISREL may be used to assess the configural invariance of the measurement model in Figure 1 across gender.

**Figure 1: measurement model**



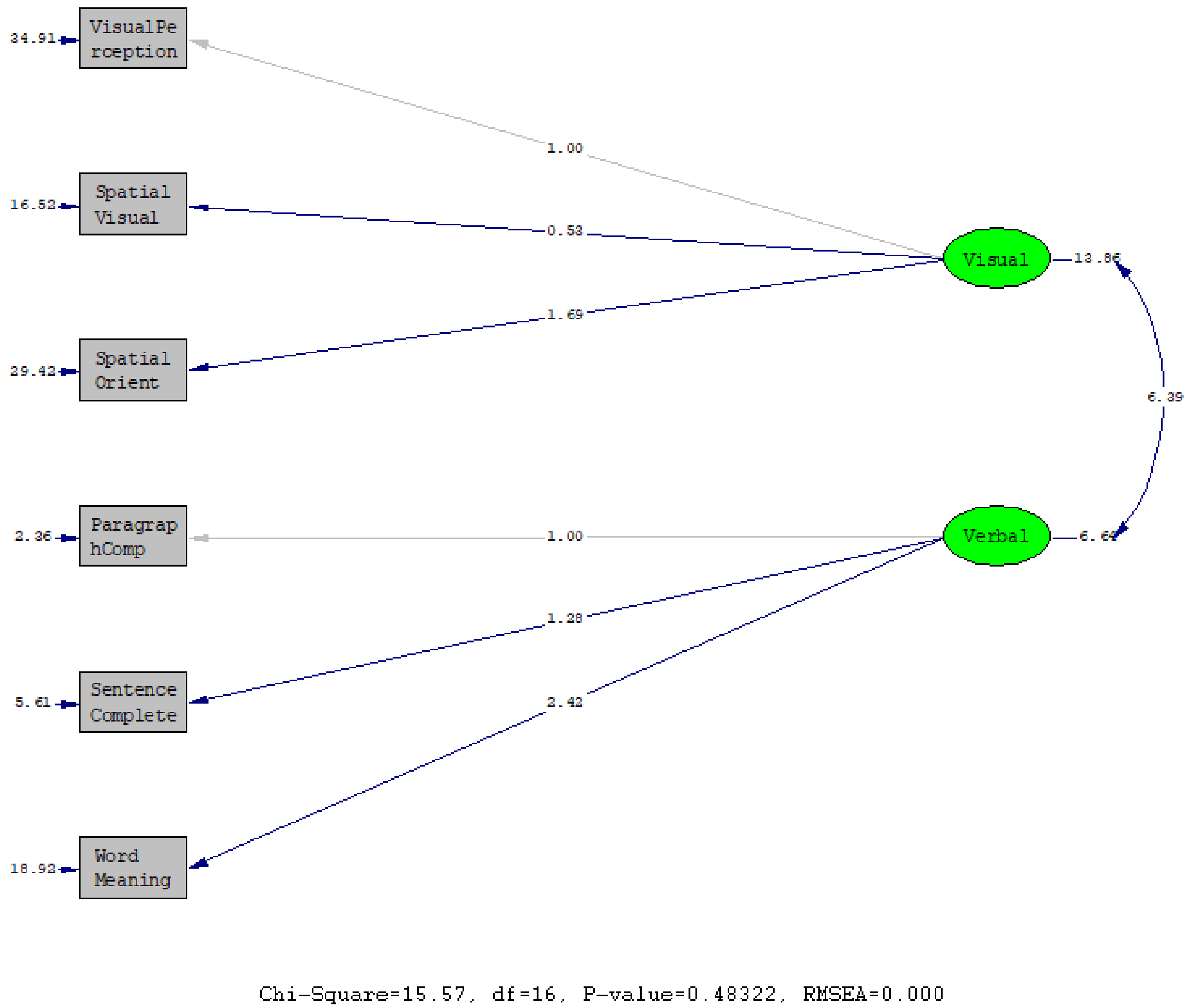

The syntax for this analysis is shown below (**LIS11_EX5.SPL**):

```
Group 1: Boys
$GROUPS=Gender
Raw Data from File LIS11_MG_BOYS_GIRLS_16.LSF
Latent Variables: Visual Verbal
VisualPerception = 1*Visual
'Spatial Visual' 'Spatial Orient' = Visual
ParagraphComp = 1*Verbal
SentenceComplete 'Word Meaning' = Verbal

Group 2: Girls
Raw Data from File LIS11_MG_BOYS_GIRLS_16.LSF
VisualPerception = 1*Visual
'Spatial Visual' 'Spatial Orient' = Visual
ParagraphComp = 1*Verbal
SentenceComplete 'Word Meaning' = Verbal
Set the Variance of Verbal Free
Set the Variance of Visual Free
Set the Covariance of Visual Verbal Free
Set the Error Variance of VisualPerception - 'Word Meaning' Free
LISREL Output: ND=3 SC
Path Diagram
End of Problem
```

Lines 11-18 specify the measurement model for boys. Lines 15-18 specify that the variance and covariance parameters of the model are different across the two groups. Line 19 requests the results in terms of the parameter matrices of the LISREL model for the measurement model in Figure 1. In addition, 3 decimal places (ND=3) and the completely standardized solutions (SC) are specified.

The path diagram and estimates obtained for this model are given below. The large *p*-value for the Chi-square test statistic value and corresponding small RMSEA value imply that the data supports the configural invariance of the measurement model in Figure 1 across boys and girls.




## 2.3.8 Adaptive quadrature example

We now use adaptive quadrature and a probit link function in this analysis based on the six ordinal variables described above. Aish & Jöreskog (1990) analyzed data on political attitudes. Their data consist of 16 ordinal variables measured on the same people at two occasions. Six of the 16 variables were considered to be indicators of political Efficacy. The attitude questions corresponding to these six variables are:

- People like me have no say in what the government does ('NOSAYINMATTERS')
- Voting is the only way that people like me can have any say about how the government runs things (VOTING)

- Sometimes politics and government seem so complicated that a person like me cannot really understand what is going on (COMPLEX)
- I don't think that public officials care much about what people like me think (NOCARE4PEOPLE)
- Generally speaking, those we elect to Parliament lose touch with the people pretty quickly (TOUCH)
- Parties are only interested in people's votes but not in their opinions (INTEREST_LEVEL)

Permitted responses to these questions were agree strongly, agree, disagree, disagree strongly, don't know and no answer.

The model fitted to the data is given in the file **efficacy2a_16.spl**.

```
Efficacy: Model 1 Estimated by FIML
Raw Data from file EFFICACY_16.LSF
$ADAPQ(8) PROBIT
Latent Variable Efficacy
Relationships
NOSAYINMATTERS - INTEREST_LEVEL = Efficacy
Path Diagram
End of Problem
End of Problem
```

Eight quadrature points are specified. Again, in order to create a new LSF file with 16-character names, we export the data from the old LSF file, amend the names as needed, and create a new LSF file. Note that the new LSF file is not downward compatible and can only be read by LISREL 11. In contrast, LSF files made by previous versions can still be opened and used in LISREL 11.

Portions of the output are given below:

```
    Measurement Equations
 NOSAYINMATTERS = 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

 NOCARE4PEOPLE = 1.656*Efficacy, Errorvar.= 1.000, R² = 0.733
 Standerr  (0.103)
 Z-values   16.007
 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_LEVEL = 1.361*Efficacy, Errorvar.= 1.000, R² = 0.649
 Standerr  (0.0744)
 Z-values   18.290
```

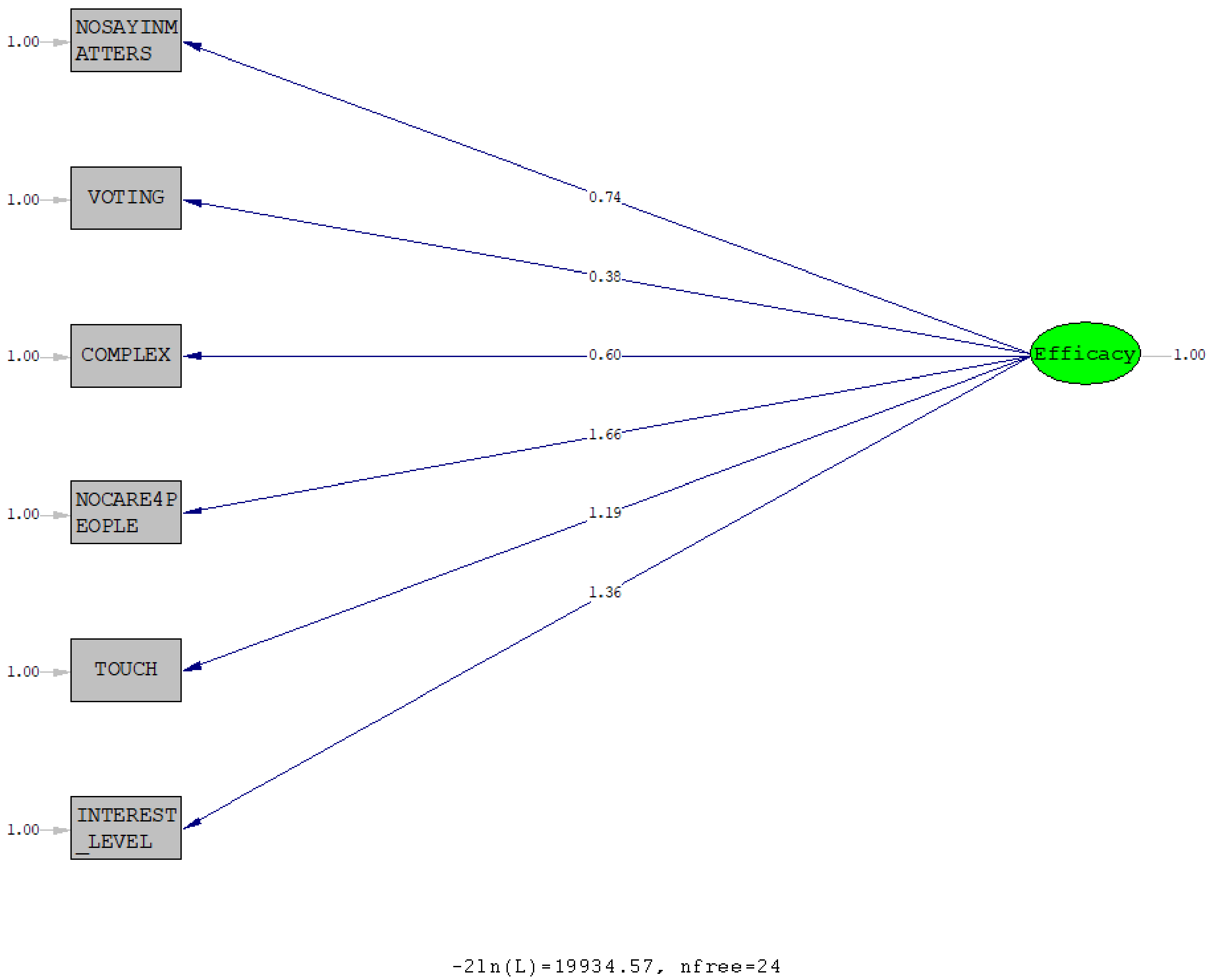
```
 P-values   0.000

Number of quadrature points =                8
     Number of free parameters =                 24
     Number of iterations used =                  7
     -2lnL (deviance statistic) =       19934.56514
     Akaike Information Criterion       19982.56514
     Schwarz Criterion                  20113.22711
```

When a cumulative log-log link function is used instead of a probit link function, the deviance statistic for that model is found to be 20069.22 with the same number of estimated parameters.

This indicates that the probit model fits the data better than the cumulative log-log model.

The following path diagram is obtained for this analysis:



## 3. Two-stage multiple imputation SEM for ordinal variables

### 3.1 Polychoric correlations

Suppose that the rows of $\mathbf{X}(n\times p)$ are $n$ observations of $p$ ordinal variables $x_1, x_2, \ldots, x_p$ with $m$ categories. Suppose further that these $p$ ordinal variables are the result of the discretization of the underlying $p$ continuous standard normal variables $z_1, z_2, \ldots, z_p$ as such that $\mathbf{z} \sim N\left(\mathbf{0}, \mathbf{P}\right)$ and

$$\begin{cases} x_i = 1 \; if \; \tau_{i0} < z_i \leq \tau_{i1} \\ x_i = 2 \; if \; \tau_{i1} < z_i \leq \tau_{i2} \\ \quad\quad\quad . \\ \quad\quad\quad . \\ \quad\quad\quad . \\ x_i = m \; if \; \tau_{i,m-1} < z_i \leq \tau_{im} \end{cases}$$

where $\mathbf{P}$ denotes the population correlation matrix of $\mathbf{z}$ and $-\infty = \tau_{i0} < \tau_{i1} < \tau_{i2} \ldots < \tau_{im} = \infty$ are parameters known as thresholds. The model for the univariate marginal of variable $x_i$ is

$$\pi_{ik} = \int_{\tau_{i,k-1}}^{\tau_{ik}} \phi(u)du$$

where $\phi\left(\cdot\right)$ denotes the probability density function of the standard normal distribution. The maximum likelihood estimator of $\tau_{ik}$ (Jöreskog, 1994) is given by

$$\hat{\tau}_{ik} = \Phi^{-1}\left(p_{i1} + p_{i2} + \cdots p_{ik}\right)$$

where $\Phi^{-1}\left(\cdot\right)$ denotes the inverse of the cumulative distribution function of the standard normal distribution and $p_{i1}, p_{i2}, \cdots, p_{im}$ denote the marginal sample proportions for $x_i$ .

The polychoric correlation matrix, $\mathbf{R}$, is a consistent estimator of the population correlation matrix $\mathbf{P}$. The model for the bivariate marginal of variables $x_i$ and $x_j$ is

$$\pi_{ijkl} = \int_{\tau_{i,k-1}}^{\tau_{ik}} \int_{\tau_{i,l-1}}^{\tau_{il}} \phi_2\left(u, v, \rho_{ij}\right) du\, dv$$

where $\phi_2\left(u, v, \rho_{ij}\right)$ denotes the probability density function of the bivariate standard normal distribution with correlation $\rho_{ij}$. The maximization of the bivariate likelihood function is equivalent to minimization of the discrepancy function

$$F\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right) = \sum_{k=1}^{m}\sum_{l=1}^{m} p_{ijkl}\left(\ln\left\{p_{ijkl}\right\} - \ln\left\{\pi_{ijkl}\right\}\right)$$

where $\hat{\boldsymbol{\tau}}_i$ and $\hat{\boldsymbol{\tau}}_j$ denote the maximum likelihood estimators of the $m-1$ thresholds of variables $x_i$ and $x_j$, respectively. The gradient of $F(\cdot)$ (Olsson (1979)) may be expressed as

$$g\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right) = -\frac{p_{ijkl}}{\pi_{ijkl}}\left[\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}\right]$$

where (Olsson (1979))

$$\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}} = \phi_2\left(\tau_{ik}, \tau_{jl}\right) - \phi_2\left(\tau_{i,k-1}, \tau_{jl}\right) + \phi_2\left(\tau_{i,k-1}, \tau_{j,l-1}\right)$$

where $\phi_2(\cdot)$ denotes the density function of the bivariate normal distribution. The information (Jöreskog, 1994) is given by

$$i\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right) = \left[\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}\right]\frac{1}{\pi_{ijkl}}\left[\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}\right]$$

The Fisher scoring algorithm is used to minimize $F(\cdot)$ with respect to $\rho_{ij}$. Let $\theta = \rho_{ij}$. If $\hat{\theta}^{(t)}$ denotes the $t^{th}$ successive approximation to $\hat{\theta}$, then the $(t+1)^{st}$ approximation is obtained from

$$\hat{\theta}^{(t+1)} = \hat{\theta}^{(t)} - \frac{g\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right)}{i\left(\rho_{ij}, \hat{\boldsymbol{\tau}}_i, \hat{\boldsymbol{\tau}}_j\right)}$$

Iteration is terminated when the absolute gradient value is below the tolerance limit $\varepsilon = 10^{-3}$.

The asymptotic covariance matrix, $\boldsymbol{\Upsilon}$, of the $p^* = p(p-1)/2$ polychoric correlations is a $p^* = p^*\left(p^*+1\right)/2$ matrix. A typical element of $\hat{\boldsymbol{\Upsilon}}$ (Jöreskog, 1994) may be expressed as

$$\left[\hat{\boldsymbol{\Upsilon}}\right]_{ij,rs} = \sum_{c=1}^{n} \kappa_{cijrs}\, \hat{\gamma}_{ijkl}\hat{\gamma}_{rsno} - \hat{\omega}_{ij}\hat{\omega}_{rs}$$

where $\kappa_{cijrs} = \dfrac{1}{n}$ if $x_{ci} = k$, $x_{cj} = l$, $x_{cr} = n$, and $x_{cs} = o$ and 0 otherwise, $\omega_{ij} = \sum_{k=1}^{m}\sum_{l=1}^{m}\gamma_{ijkl}\pi_{ijkl}$, $\omega_{rs} = \sum_{n=1}^{m}\sum_{o=1}^{m}\gamma_{rsno}\pi_{rsno}$, and $\gamma_{ijkl}$ denotes a typical element of

$$\boldsymbol{\Gamma}_{ij} = \boldsymbol{\alpha}_{ij} + \mathbf{B}_i\boldsymbol{\beta}_i\mathbf{1}_j' + \mathbf{1}_j\boldsymbol{\beta}_j'\mathbf{B}_j'$$

where $\mathbf{1}_i$ denotes an $m\times 1$ column vector and

$$\mathbf{B}_i' = \left(\mathbf{A}_i'\mathbf{D}_{\pi_i}^{-1}\mathbf{A}_i\right)^{-1}\mathbf{A}_i'\mathbf{D}_{\pi i_j}^{-1}$$

$$\mathbf{B}_j' = \left(\mathbf{A}_j'\mathbf{D}_{\pi_j}^{-1}\mathbf{A}_j\right)^{-1}\mathbf{A}_j'\mathbf{D}_{\pi_j}^{-1}$$

Typical elements of $\boldsymbol{\alpha}_{ij}$, $\boldsymbol{\beta}_i$, and $\boldsymbol{\beta}_j$ are given by

$$\alpha_{ijkl} = D^{-1}\frac{1}{\pi_{ijkl}}\frac{\partial \pi_{ijkl}}{\rho_{ij}}$$

$$\left[\boldsymbol{\beta}_i\right]_k = \sum\nolimits_{k=1}^{m}\sum\nolimits_{l=1}^{m}\alpha_{ijkl}\left(\frac{\partial \pi_{ijkl}}{\tau_{ik}}\right)$$

$$\left[\boldsymbol{\beta}_j\right]_l = \sum\nolimits_{k=1}^{m}\sum\nolimits_{l=1}^{m}\alpha_{ijkl}\left(\frac{\partial \pi_{ijkl}}{\tau_{il}}\right)$$

where

$$D = \sum\nolimits_{k=1}^{m}\sum\nolimits_{l=1}^{m}\frac{1}{\pi_{ijkl}}\left(\frac{\partial \pi_{ijkl}}{\partial \rho_{ij}}\right)^2$$

Structural equation models for ordinal variables can be fitted to the polychoric correlation matrix and the estimated asymptotic covariance matrix of the polychoric correlations by using the robust DWLS, WLS, or ULS methods (Chung and Cai (2019)).

## 3.2 Multiple imputation

Suppose now that the $n$ observations of the $p$ ordinal variables include missing data values with $k$ missing data value patterns. The EM algorithm and the MMMC method for multiple imputation of incomplete data are intended for continuous variables and cannot readily be applied to ordinal variables. However, they can be applied to the underlying continuous variables $z_1, z_2, \ldots, z_p$ associated with the ordinal variables $x_1, x_2, \ldots, x_p$. Although no observations for these continuous variables are available, these variables are assumed to have a multivariate standard normal distribution with a population covariance matrix $\boldsymbol{\Sigma}$. As a result, we can simulate data from this distribution by using the polychoric correlation matrix of the complete data of the variables if the number of complete cases is large enough and use either the EM algorithm or the MCMC algorithm to impute the missing data values for the underlying continuous variables. After imputation, the estimated thresholds can be used to replace the missing data values for the corresponding ordinal variables.

Suppose that the rows of $\mathbf{Z}(n \times p)$ are $n$ observations of the $p$ underlying continuous variables $z_1, z_2, \ldots, z_p$ simulated from the $N\left(\mathbf{0}, \boldsymbol{\Sigma}\right)$ distribution and that $\mathbf{Z}_o$ denotes the observed data values that corresponds with the observed data values of $\mathbf{X}$. The EM algorithm (Dempster, Laird, and Rubin 1977) can be used to compute the maximum likelihood estimate of $\boldsymbol{\Sigma}$. The minus two observed-data log likelihood may be expressed as

$$-2\ln L(\boldsymbol{\Sigma} \mid \mathbf{Z}_o) = \sum_{i=1}^{k} n_i \ln\left|\boldsymbol{\Sigma}_i\right| + \sum_{i=1}^{k}\sum_{j=1}^{n_i}\mathbf{z}_{oij}^{'}\boldsymbol{\Sigma}_i^{-1}\,1\,\mathbf{z}_{oij}$$

where $n_i$ denotes the number of observations of missing data value pattern $i = 1, 2, \ldots, k$, $\boldsymbol{\Sigma}_i$ denotes the population covariance matrix for missing data value pattern $i$, and $\mathbf{z}_{oij}$ is the $j^{th}$ vector of observed values of missing data value pattern $i$.

The initial estimate for the M-step is the sample covariance matrix, $\mathbf{S}$, of the complete data or $\mathbf{I}_p$ if the number of complete observations is too small. In the E-step, the conditional covariance matrices of the missing variables given the observed variables for the missing data value patterns are computed and used to compute an updated estimate $\hat{\mathbf{\Sigma}}^{(t+1)}$ of $\mathbf{\Sigma}$. Iteration of the consecutive M and E steps is terminated when the absolute difference between $\hat{\mathbf{\Sigma}}^{(t+1)}$ and $\hat{\mathbf{\Sigma}}^{(t)}$ is below the tolerance limit $\varepsilon = 10^{-5}$.

The correlation matrix of the EM estimate, $\hat{\mathbf{\Sigma}}$, of $\mathbf{\Sigma}$ is used as the initial covariance matrix of the multivariate standard normal distribution in the first step of the Monte Carlo Markov Chain (MCMC) method. In the first step ($\mathbf{\Sigma}$-step) of the MCMC method, an estimate of $\mathbf{\Sigma}$ is simulated form an inverse Wishart distribution. In the I-step, observations are simulated from the conditional standard normal distributions of the missing variables given the observed $k$ missing data value patterns and used to replace the missing data values. The next estimate of $\mathbf{\Sigma}$ is then obtained by computing the sample correlation matrix of the imputed data. The P and I steps are repeated for a fixed number of times.

Let the rows of $\mathbf{Z}_i(n \times p)$ contain the observed and imputed data values for the standard normal variables $z_1, z_2, \ldots, z_p$. The observed data for the ordinal variables are obtained from the corresponding observed data values of $\mathbf{X}$. The missing data values of $\mathbf{X}$ are then replaced by the values obtained from using the corresponding imputed data values of $\mathbf{Z}$ and the estimated thresholds.

## 3.3 Average moment matrices

Suppose that $\mathbf{X}_{1i}, \mathbf{X}_{2i}, \ldots, \mathbf{X}_{mi}$ are $m$ MCMC imputed data sets for the incomplete data matrix, $\mathbf{X}$, of the $p$ ordinal variables $x_1, x_2, \ldots, x_p$ and that $\hat{\mathbf{P}}_1, \hat{\mathbf{P}}_2, \ldots, \hat{\mathbf{P}}_m$ and $\hat{\mathbf{\Upsilon}}_1, \hat{\mathbf{\Upsilon}}_2, \cdots, \hat{\mathbf{\Upsilon}}_m$ denote the corresponding polychoric correlation matrices and the estimated asymptotic covariance matrices of the polychoric correlations, respectively. Then the average polychoric correlation matrix is

$$\bar{\mathbf{P}} = \frac{1}{m}\sum_{i=1}^{m}\hat{\mathbf{P}}_i$$

and the average estimated asymptotic covariance matrix is

$$\bar{\mathbf{\Upsilon}} = \frac{1}{m}\sum_{i=1}^{m}\hat{\mathbf{\Upsilon}}_i$$

Chung and Cai (2019) point out that $\bar{\mathbf{\Upsilon}}$ only captures uncertainty based on complete data. As a result, its inverse cannot be used as a weight matrix for the robust DWLS, WLS, and ULS methods for ordinal structural equational modeling. A corrected weight matrix is obtained by correcting for the between-imputation variation in the estimated polychoric correlations and is obtained as the inverse of

$$\tilde{\mathbf{\Upsilon}} = \bar{\mathbf{\Upsilon}} + \frac{m+1}{m(m-1)}\left[\sum_{i=1}^{m}(\hat{\boldsymbol{\rho}}_i - \bar{\boldsymbol{\rho}})(\hat{\boldsymbol{\rho}}_i - \bar{\boldsymbol{\rho}})'\right]$$

where $\mathbf{s}$ denotes the $p \times (p-1)/2$ vector consisting of the nonduplicated elements of the $p \times p$ symmetric matrix $\mathbf{S}$. $\bar{\mathbf{P}}$ and $\bar{\mathbf{\Upsilon}}$ can be used to fit structural equation models to the average polychoric correlation matrix with the robust DWLS, WLS, and ULS methods. The corrected robust DWLS and ULS Chi-square test statistic proposed by Chung and Cai (2019) is given by

$$T_B = (n-1)\mathbf{r}' \tilde{\mathbf{\Omega}} \mathbf{r}$$

where $\mathbf{r} = \bar{\boldsymbol{\rho}} - \boldsymbol{\rho}(\boldsymbol{\theta})$ and

$$\tilde{\mathbf{\Omega}} = \tilde{\mathbf{\Upsilon}}^{-1} - \tilde{\mathbf{\Upsilon}}^{-1}\tilde{\mathbf{\Delta}}\left(\tilde{\mathbf{\Delta}}'\tilde{\mathbf{\Delta}}\right)^{-1}\tilde{\mathbf{\Delta}}'\tilde{\mathbf{\Upsilon}}^{-1}$$

where $\tilde{\mathbf{\Delta}}$ denotes the Jacobian matrix of $\boldsymbol{\rho}$ with respect to the unknown parameters $\boldsymbol{\theta}$ of the structural equation model evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. The small sample adjusted $T_B$ test statistic is given by

$$T_{YB} = \frac{T_B}{1 + nT_B / (n-1)}$$

## 3.4 Goodness-of-fit statistics

### 3.4.1 Chi-square test statistic values

The Chi-square test statistic values computed in LISREL are based on the asymptotical statistical theory in Browne (1974, 1984, 1987), Jöreskog (1981), and Satorra and Bentler (1988, 1994, 2001, 2010). The exact formulae for computing the Maximum Likelihood Ratio Chi-Square ($c_1$), the Browne's (1984) ADF Chi-Square ($c_2$ with weight matrix $\mathbf{W}_{NT}$), the Satorra-Bentler (1988) Scaled Chi-Square ($c_3$), and the Satorra-Bentler (1988) Adjusted Chi-Square ($c_4$) are provided in Jöreskog et. al. (2001).

LISREL now computes two different values for $c_2$. The first value, C2_NT, is based on the $c_2$ formula provided in Jöreskog et al (2001) while the second value, C2_NNT, is obtained by replacing the weight matrix $\mathbf{W}_{NT}$ with $\mathbf{W}_{NNT}$. LISREL also computes the Chi-Square Scaled and Shifted ($c_5$) and the Yuan-Bentler (1997) Chi-Square ($c_6$). The formula for $c_5$ is provided in Satorra and Bentler (2010). The value of $c_6$ is computed as

$$c_6 = \frac{c_s}{1 + nc_s / (n-1)}$$

where $n$ denotes the sample size and $c_s = c_1$ for robust Weighted Least Squares (WLS) estimation and $c_s = c_2$ based on $\mathbf{W}_{NNT}$ for robust Diagonally Weighted Least Squares (DWLS) estimation and robust Unweighted Least Squares (ULS) estimation.

### 3.4.2 Root Mean Square Error of Approximation (RMSEA) estimates

A point estimate and an interval estimate of the Root Mean Square Error of Approximation (RMSEA) are provided in Steiger and Lind (1981). These estimates depend on the Chi-square test statistic value. The point estimate of the RMSEA may be expressed as

$$\hat{r}_p = \sqrt{\frac{\max\{c_r - d, 0\}}{(n-1)d}}$$

where $c_r$ denotes the Chi-square test statistic value for $d$ degrees of freedom. The 90% confidence interval estimate for the RMSEA is given by

$$\hat{r}_i = \left( \sqrt{\frac{\lambda_l}{(n-1)d}}; \sqrt{\frac{\lambda_u}{(n-1)d}} \right)$$

where $\lambda_l$ and $\lambda_u$ denotes the 5th and the 95th percentiles of a non-central Chi-square distribution with d degrees of freedom with non-centrality parameter, respectively.

LISREL uses $c_r = c_1$ for Maximum Likelihood (ML), Generalized Least Squares (GLS), and WLS estimation, $c_r = c_4$ for robust ML estimation, and $c_r = c_2$ based on $\mathbf{W}_{NNT}$ for robust GLS, robust ULS, and robust DWLS estimation.

## 3.5 Syntax

The new syntax for the two-stage multiple imputation SEM method consists of two new keywords for the LISREL OU command and the LISREL Output SIMPLIS command.

## 3.5.1 OU command in LISREL

**Purpose**

To specify the estimation methods to be used and to specify the results to be produced. For keywords and options on this command not related to the MI2S routine, please refer to the LISREL syntax guide.

**Status**

Required.

**Syntax**

```
OU <keywords> <options>
```

The existing keywords and options for the OU command are described in the LISREL Syntax Guide which is available via the Help menu. In this new MI2S implementation, the ME keyword is required to specify the method of estimation and only the DWLS, ULS, and DWLS options are permitted for two-stage multiple imputation SEM (MI2S).

### MI2S option

**Purpose**

To specify the two-stage multiple imputation structural equation modeling method for ordinal variables.

**Status**

Optional, unless the two-stage multiple imputation structural equation modeling method for ordinal variables is desired.

**Syntax**

```
MI2S
```

### NM Keyword

#### Purpose

To specify the number of multiple MCMC imputations.

#### Status

Optional.

#### Syntax

```
NM = <number>
```

where <number> denotes a positive integer greater than zero.

#### Example

```
NM = 50
```

#### Default

```
NM = 10
```

### IX Keyword

#### Purpose

To specify the integer starting value (random seed) for the random number generator.

#### Status

Optional, unless a specific starting random seed is desired.

#### Syntax

```
IX=<number>
```

where <number> denotes a positive integer greater than zero.

#### Example

```
IX=4087
```

#### Default

```
IX=123456
```

## 3.5.2 LISREL Output command in SIMPLIS

#### Purpose

To request the results to be printed in terms of the specific LISREL model, to specify the methods to be used, and to specify the results to be produced. For keywords and options on this command not related to the MI2S routine, please refer to the SIMPLIS syntax guide.

#### Status

Optional, unless the two-stage multiple imputation SEM method is desired.

### Syntax

```
LISREL Output <keywords> <options>
```

The existing keywords and options for the LISREL Output command are described in the SIMPLIS Syntax Guide which is available via the Help menu. In this new MI2S implementation, the ME keyword is required and only the DWLS, ULS, and DWLS options are permitted for two-stage multiple imputation SEM.

## MI2S option

### Purpose

To specify the two-stage multiple imputation structural equation modeling method for ordinal variables.

### Status

Optional, unless the two-stage multiple imputation structural equation modeling method for ordinal variables is desired.

### Syntax

```
MI2S
```

## NM Keyword

### Purpose

To specify the number of multiple MCMC imputations.

### Status

Optional, unless the two-stage multiple imputation SEM method is desired.

### Syntax

```
NM = <number>
```

where <number> denotes a positive integer greater than zero.

### Example

```
NM = 50
```

### Default

```
NM = 10
```

## IX Keyword

### Purpose

To specify the integer starting value (random seed) for the random number generator.

### Status

Optional, unless a specific starting random seed is desired.

### Syntax

```
IX=<number>
```

where <number> denotes a positive integer greater than zero.

**Example**

```
IX=4087
```

**Default**

```
IX=123456
```

## 3.6 Two-stage SEM examples

### 3.6.1 Political efficacy data

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. The first few lines of this data set are shown below.

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 |

Note that the data values of -999999.00 are missing data values. Should a different code be used to indicate missing values, it should be assigned as the global missing code using the **Define Variables** dialog box.

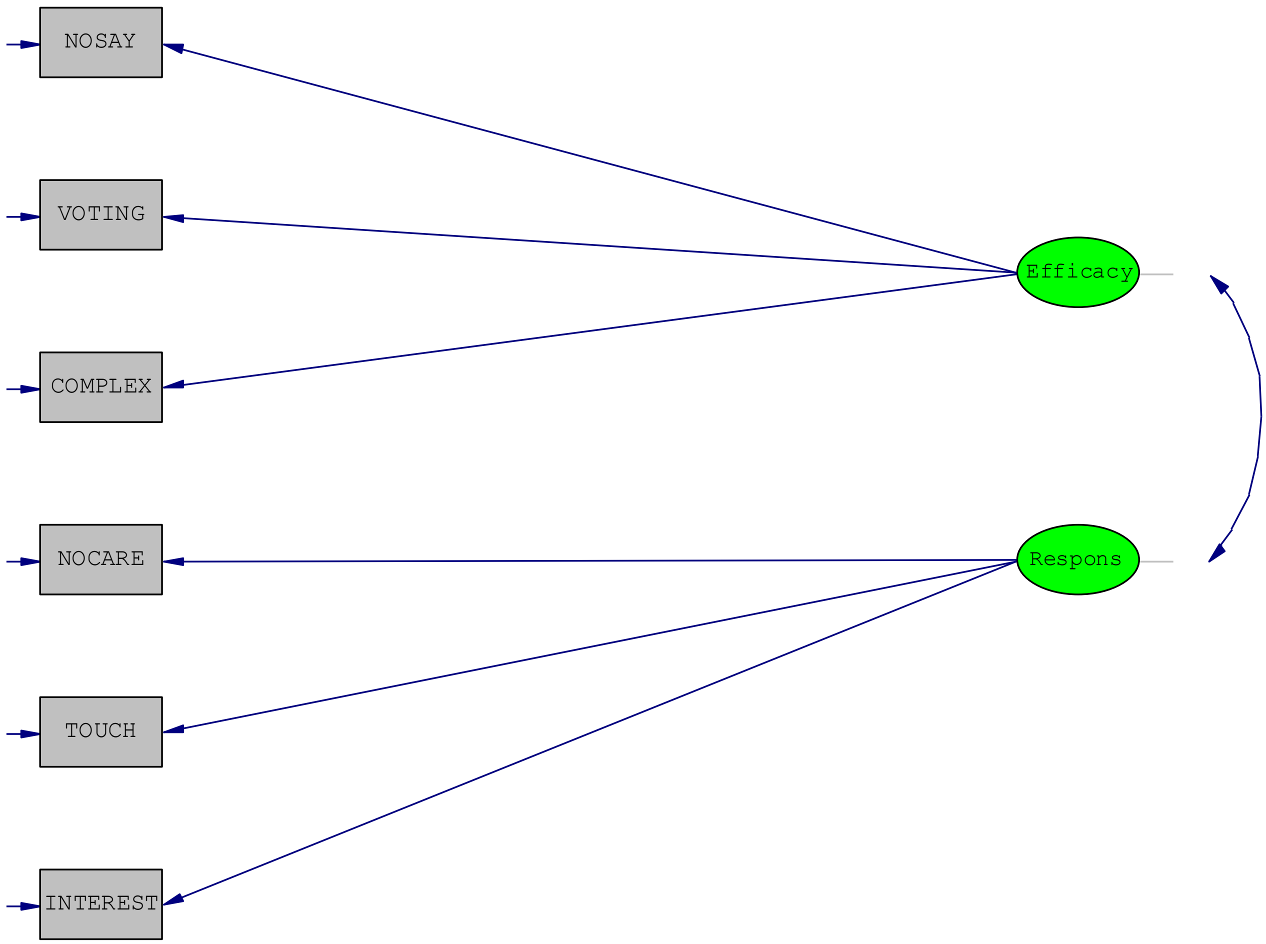
The data are the responses to the following statements:

- People like me have no say in what the government does (NOSAY)
- Voting is the only way that people like me can have any say about how the government runs things (VOTING)
- Sometimes politics and government seem so complicated that a person like me cannot really understand what is going on (COMPLEX)
- I don't think that public officials care much about what people like me think (NOCARE)
- Generally speaking, those we elect to Parliament lose touch with the people pretty quickly (TOUCH)
- Parties are only interested in people's votes but not in their opinions (INTEREST)

The ordered categories are:

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

The theoretical measurement model is a confirmatory factor analysis model that specifies that the six political variables are indicators of political efficacy and political responsiveness. A path diagram of the theoretical model is shown in the image below.




The SIMPLIS syntax file to fit the model reflected in the path diagram above to the average polychoric correlation matrix for 10 multiple MCMC imputations is depicted in the image below.

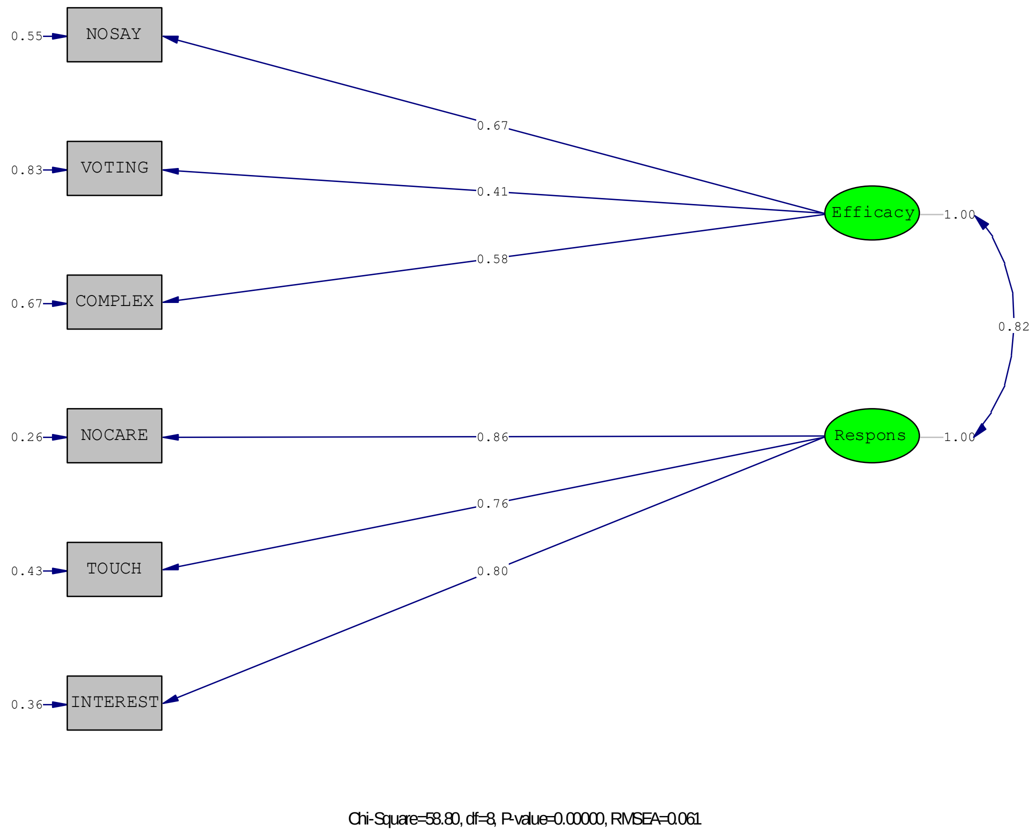
EFFICACY1A.SPL

```
Raw Data from File EFFICACY.LSF
Latent Variables
Efficacy Respons
Relationships
NOSAY VOTING COMPLEX = Efficacy
NOCARE TOUCH INTEREST = Respons
LISREL Output: ND=3 SC MI2S IX=37996 NM=10 ME=DWLS
Path Diagram
End of Problem
```

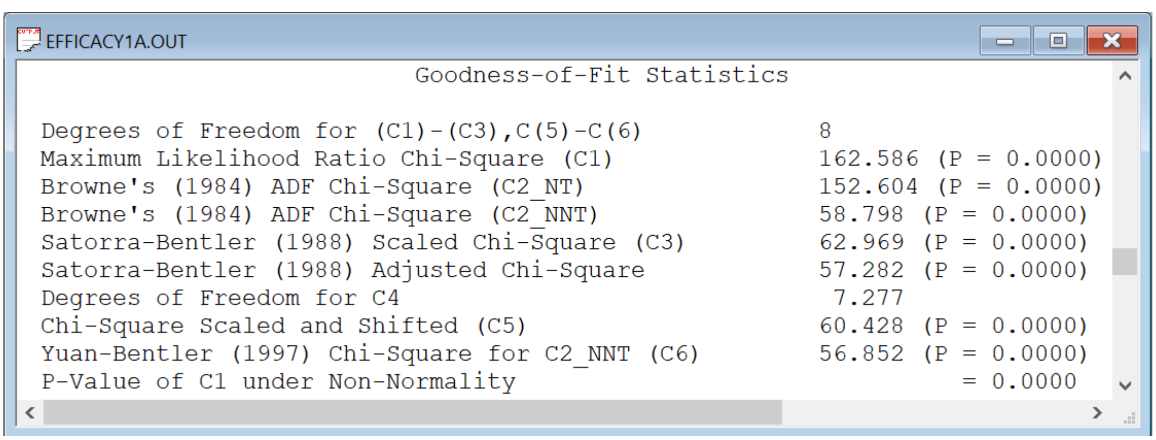
- Line 1 specifies the raw data source.
- Lines 2 and 3 specify labels for the latent variables of the model.
- Lines 4 to 6 specify the measurement model for the latent variables political efficacy and political responsiveness.
- Line 7 requests that the results in the output file should be given in terms of the LISREL model for the structural equation model (LISREL Output). It also requests that the results should be written to three decimal places (ND= 3), that the completely standardized solution should be written to the output file (SC), and robust diagonally weighted least squares estimation (ME = DWLS).

- The MI2S option invokes the two-stage multiple imputation SEM method to fit the model to the average polychoric correlation matrix for the NM = 10 MCMC imputations based on an initial random seed of IX = 37966.
- Line 8 requests a path diagram of the model.
- Line 9 indicates that no more SIMPLIS commands are to be processed.

If the above **SPL** file is opened with LISREL and the **Run LISREL** button is clicked, the following path diagram is obtained.




The corresponding output file, **EFFICACY1A.OUT**, is opened in a separate window. A small portion of this file is shown in the following image.



EFFICACY1A.OUT

```
                     Goodness-of-Fit Statistics

 Degrees of Freedom for (C1)-(C3),C(5)-C(6)            8
 Maximum Likelihood Ratio Chi-Square (C1)              162.586 (P = 0.0000)
 Browne's (1984) ADF Chi-Square (C2_NT)                152.604 (P = 0.0000)
 Browne's (1984) ADF Chi-Square (C2_NNT)               58.798 (P = 0.0000)
 Satorra-Bentler (1988) Scaled Chi-Square (C3)         62.969 (P = 0.0000)
 Satorra-Bentler (1988) Adjusted Chi-Square            57.282 (P = 0.0000)
 Degrees of Freedom for C4                              7.277
 Chi-Square Scaled and Shifted (C5)                    60.428 (P = 0.0000)
 Yuan-Bentler (1997) Chi-Square for C2_NNT (C6)        56.852 (P = 0.0000)
 P-Value of C1 under Non-Normality                              = 0.0000
```

The Chi-square test statistic values above indicate that the theoretical measurement model for political efficacy and political responsiveness is not supported by the data.

### 3.6.2 Attitudes of morality and equality

In this example, we return to the data discussed in Section 2.3.2, namely study of attitudes of morality and equality (Hasselrot & Lernberg (1980)), Swedish school children in grade 9 were asked questions about their attitudes regarding social issues in family, school, and society. Among the questions asked were the following eight items.

For me, questions about……

1. human rights
2. equal conditions for all people
3. racial problems
4. equal value of all people
5. euthanasia
6. crime and punishment
7. conscientious objectors
8. guild and bad conscience

are….

unimportant    not important    important    very important

For the present example we use a subsample of 200 cases. Responses to the eight questions were scored 1, 2, 3, and 4, where 4 denotes very important. The data matrix consists of 200 rows and 8 columns and is stored in the file **MORALITY.LSF**.

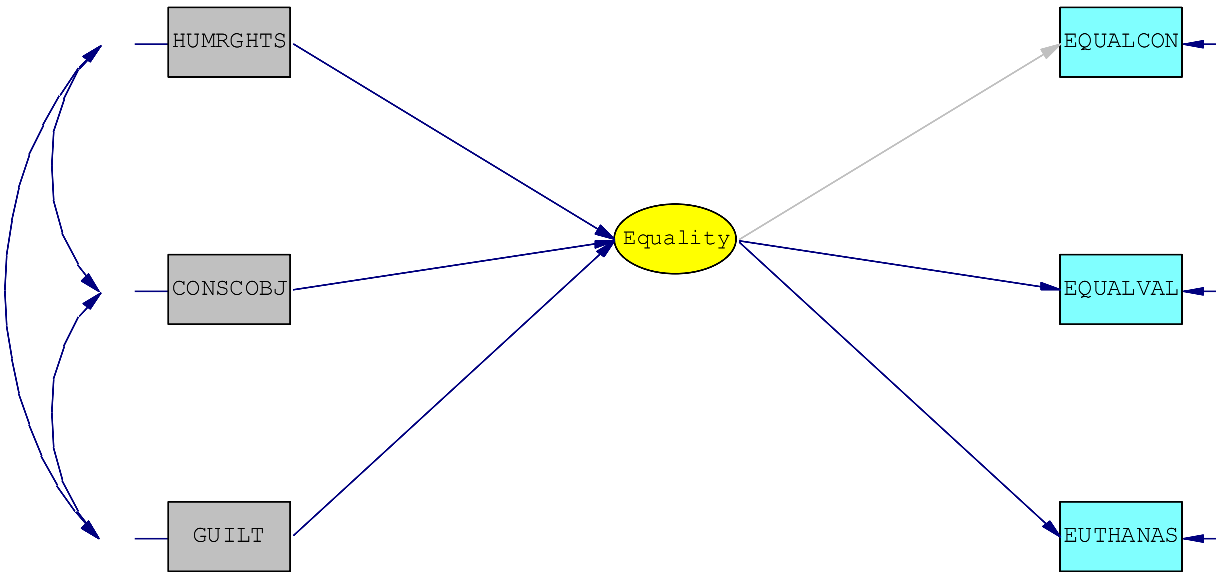
The first few cases of this data file are reflected in the image below.

MORALITY.lsf

| | HUMRGHTS | EQUALCON | RACEPROB | EQUALVAL | EUTHANAS | CRIMEPUN | CONSCOBJ | GUILT |
|---|---|---|---|---|---|---|---|---|
| 1 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 3.00 | 4.00 |
| 2 | 4.00 | 3.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 1.00 |
| 3 | 4.00 | 4.00 | 4.00 | 3.00 | 4.00 | 3.00 | 3.00 | 3.00 |
| 4 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 3.00 | 4.00 |
| 5 | -999999.00 | -999999.00 | -999999.00 | 4.00 | 4.00 | 4.00 | 3.00 | 3.00 |
| 6 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| 7 | 3.00 | 2.00 | 4.00 | 3.00 | 3.00 | 3.00 | 2.00 | 3.00 |
| 8 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 | 3.00 |
| 9 | 3.00 | 4.00 | 4.00 | 4.00 | 4.00 | 3.00 | 4.00 | 4.00 |
| 10 | 3.00 | 3.00 | 3.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 |
| 11 | 3.00 | 3.00 | 3.00 | 4.00 | 3.00 | 4.00 | 1.00 | 3.00 |
| 12 | 4.00 | 3.00 | 4.00 | 3.00 | 2.00 | 4.00 | 3.00 | 3.00 |
| 13 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 | 4.00 |
| 14 | 4.00 | 3.00 | 2.00 | 3.00 | 3.00 | 3.00 | 2.00 | 2.00 |

Note that the data values of -999999.00 are missing data values. Should a different code be used to indicate missing values, it should be assigned as the global missing code using the **Define Variables** dialog box.

The theoretical model is a MIMIC model that specifies that three attitude variables are reflective indicators of equality, and three attitude variables are formative indicators of equality as reflected in the path diagram shown in the image below.




The SIMPLIS syntax file to fit the model reflected in the path diagram to the average polychoric correlation matrix for 50 multiple MCMC imputations is depicted in the image below. The two-stage multiple imputation SEM syntax is reflected on the LISREL Output command as MI2S which invokes the method, NM = 50 which specifies 50 multiple imputations, and IX = 4816 specifies a starting random seed of 4816.

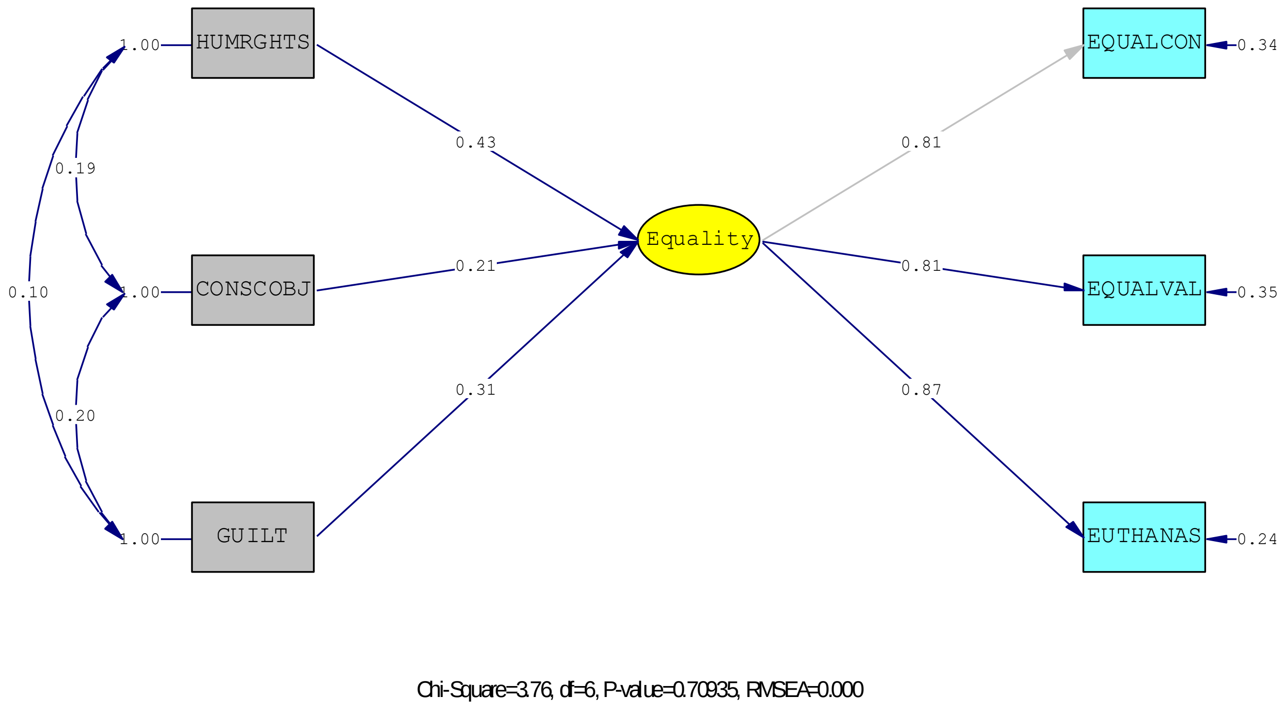
MORALITY2A.SPL

```
Raw Data from File MORALITY.LSF
Latent Variables
Equality
Relationships
EQUALCON EQUALVAL EUTHANAS = Equality
Equality = HUMRGHTS CONSCOBJ GUILT
LISREL Output: ND=3 SC MI2S ME=WLS IX=4816 NM=50
Path Diagram
End of Problem
```

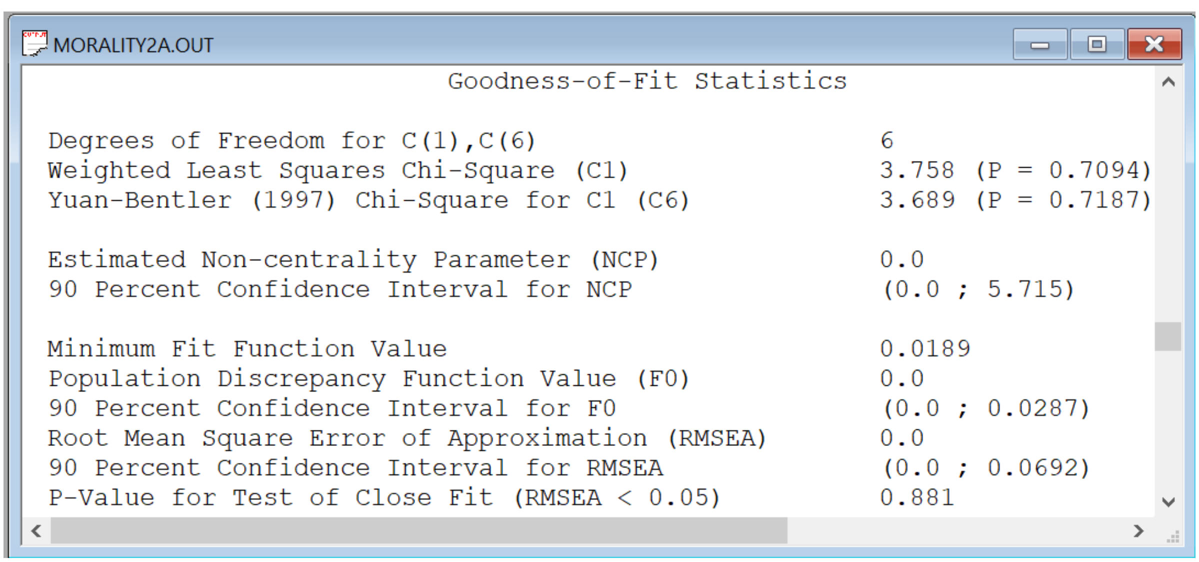
- Line 1 specifies the raw data source.
- Lines 2 and 3 specify labels for the latent variable of the model.
- Lines 4 to 6 specify the MIMIC model equality.
- Line 7 requests that the results in the output file should be given in terms of the LISREL model for the structural equation model (LISREL Output). It also requests that the results should be written to three decimal places (ND = 3), that the completely standardized solution should be written to the output file (SC), and robust weighted least squares estimation (ME = WLS).
- The MI2S option invokes the two-stage multiple imputation method to fit the model to the average polychoric correlation matrix for the NM = 50 MCMC imputations based on an initial random seed of IX = 4816.
- Line 8 requests a path diagram of the model.
- Line 9 indicates that no more SIMPLIS commands are to be processed.

### The results

If the above **SPL** file is opened with LISREL and the **Run LISREL** button is clicked, the following path diagram is obtained.




The corresponding output file, **MORALITY2A.OUT**, is opened in a separate window. A small portion of this file is shown in the following image.


MORALITY2A.OUT

```
                              Goodness-of-Fit Statistics

 Degrees of Freedom for C(1),C(6)                            6
 Weighted Least Squares Chi-Square (C1)                      3.758 (P = 0.7094)
 Yuan-Bentler (1997) Chi-Square for C1 (C6)                  3.689 (P = 0.7187)

 Estimated Non-centrality Parameter (NCP)                    0.0
 90 Percent Confidence Interval for NCP                      (0.0 ; 5.715)

 Minimum Fit Function Value                                  0.0189
 Population Discrepancy Function Value (F0)                  0.0
 90 Percent Confidence Interval for F0                       (0.0 ; 0.0287)
 Root Mean Square Error of Approximation (RMSEA)             0.0
 90 Percent Confidence Interval for RMSEA                    (0.0 ; 0.0692)
 P-Value for Test of Close Fit (RMSEA < 0.05)                0.881
```

The values of the goodness-of-fit statistics above indicate that the theoretical model for equality is supported by the data.

## 3.6.3 Political efficacy panel data

This example is based on panel data of the six political efficacy measurements as described in Aish & Jöreskog (1990) observed in two different calendar years. The dataset **PANELUSA.LSF** consists of 933 cases obtained in a USA sample. The first few lines of this data set are shown below.

panelusa.LSF

| | NOSAY1 | VOTING1 | COMPLEX1 | NOCARE1 | TOUCH1 | INTERES1 | NOSAY2 | VOTING2 |
|---|---|---|---|---|---|---|---|---|
| 1 | 2.00 | 2.00 | 1.00 | 1.00 | 1.00 | 1.00 | -999999.00 | 2.00 |
| 2 | 2.00 | 3.00 | 3.00 | 3.00 | 2.00 | 3.00 | 2.00 | 3.00 |
| 3 | 3.00 | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 |
| 4 | 2.00 | 2.00 | 1.00 | 1.00 | 2.00 | 1.00 | 2.00 | 2.00 |
| 5 | 3.00 | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 | 2.00 |
| 6 | 2.00 | 2.00 | 2.00 | 2.00 | 1.00 | 2.00 | 3.00 | 2.00 |
| 7 | 3.00 | 1.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 | 2.00 |
| 8 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 | 3.00 | 3.00 |
| 9 | 3.00 | 3.00 | 2.00 | 2.00 | 3.00 | 3.00 | 3.00 | 3.00 |
| 10 | 2.00 | 2.00 | 3.00 | 1.00 | 1.00 | 1.00 | 2.00 | 2.00 |
| 11 | 3.00 | 2.00 | 1.00 | 1.00 | 2.00 | 2.00 | 3.00 | 2.00 |
| 12 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 3.00 | 3.00 |
| 13 | 2.00 | 2.00 | 2.00 | 1.00 | 2.00 | 2.00 | 1.00 | 1.00 |
| 14 | 3.00 | 3.00 | 2.00 | 3.00 | 2.00 | 2.00 | 3.00 | 2.00 |

Note that the data values of -999999.00 are missing data values. Should a different code be used to indicate missing values, it should be assigned as the global missing code using the **Define Variables** dialog box.

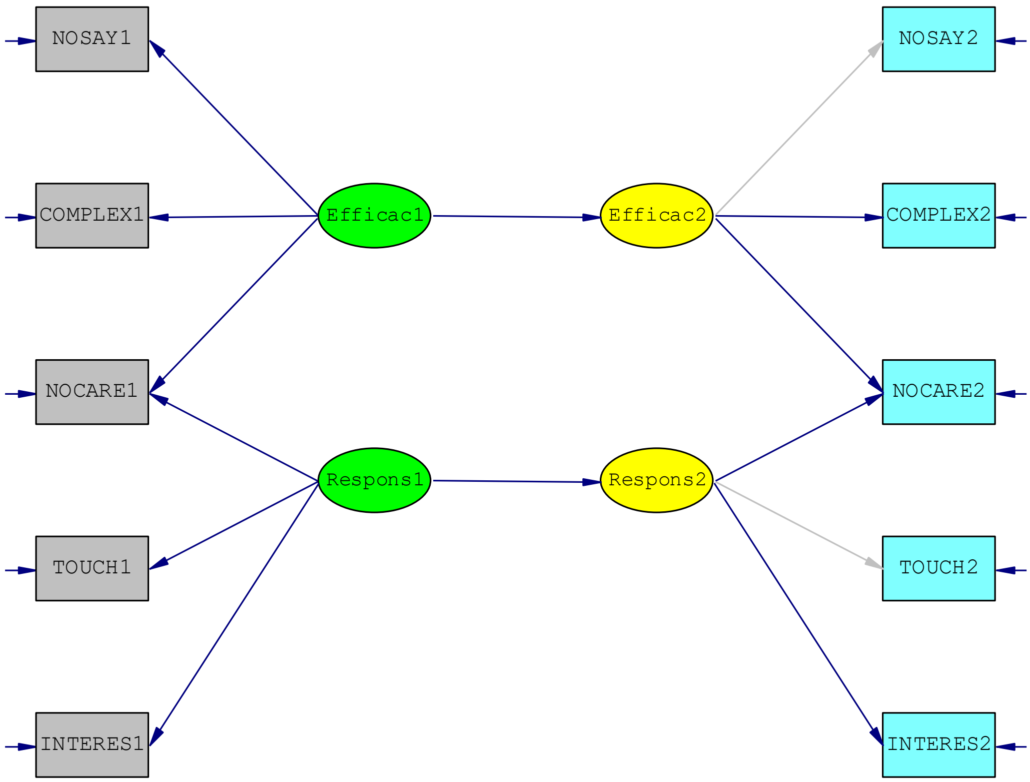
The data are the responses to the following statements:

- People like me have no say in what the government does (NOSAY)
- Voting is the only way that people like me can have any say about how the government runs things (VOTING)
- Sometimes politics and government seem so complicated that a person like me cannot really understand what is going on (COMPLEX)
- I don’t think that public officials care much about what people like me think (NOCARE)
- Generally speaking, those we elect to Parliament lose touch with the people pretty quickly (TOUCH)
- Parties are only interested in people’s votes but not in their opinions (INTEREST)

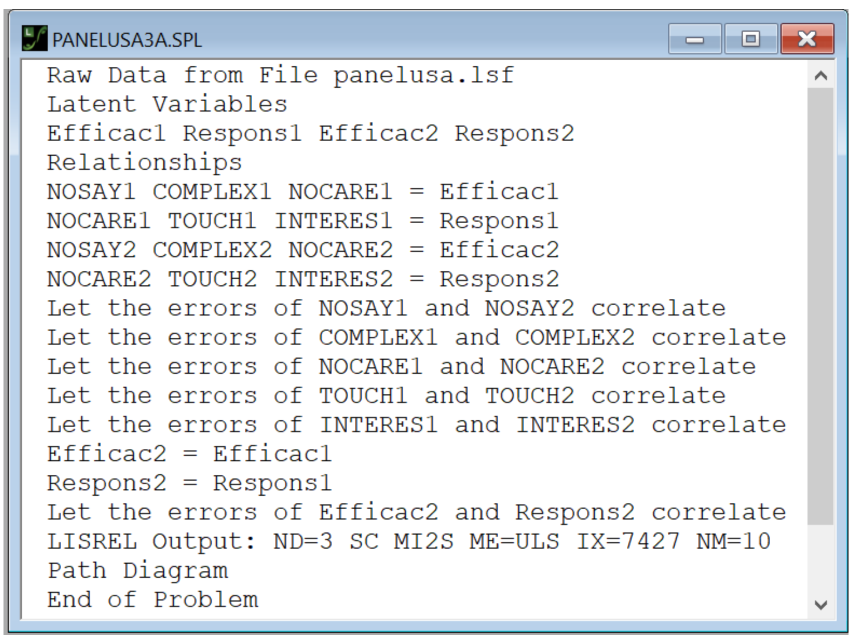
The ordered categories are:

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

The theoretical model is a two-wave model for political efficacy and political responsiveness. A path diagram of the theoretical model is shown in the image below.




The SIMPLIS syntax file to fit the model reflected in the path diagram above to the average polychoric correlation matrix for 10 multiple MCMC imputations is depicted in the image below. The two-stage multiple imputation SEM syntax is reflected on the LISREL Output command as MI2S which requests the method, NM = 10 which requests 10 multiple imputations, and IX = 7427 which requests a starting random seed of 7427.

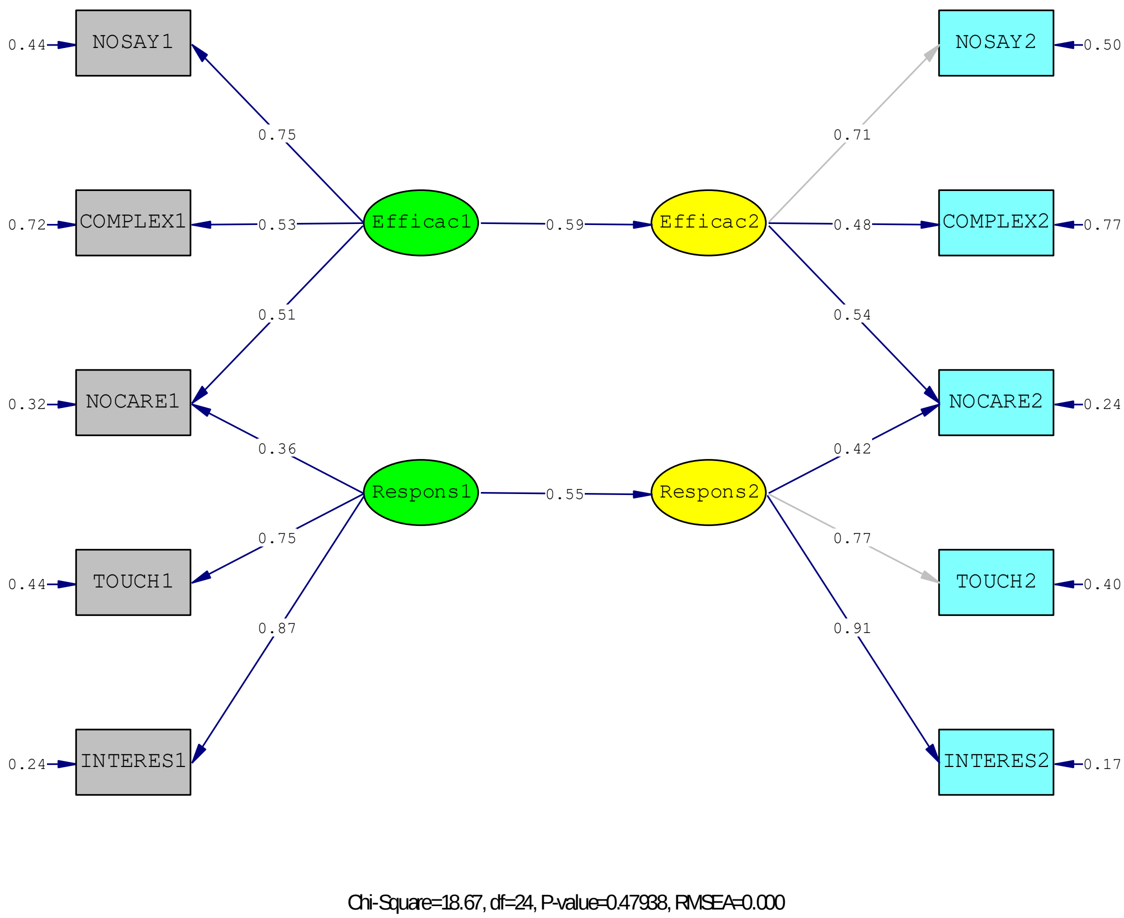

PANELUSA3A.SPL

```
Raw Data from File panelusa.lsf
Latent Variables
Efficac1 Respons1 Efficac2 Respons2
Relationships
NOSAY1 COMPLEX1 NOCARE1 = Efficac1
NOCARE1 TOUCH1 INTERES1 = Respons1
NOSAY2 COMPLEX2 NOCARE2 = Efficac2
NOCARE2 TOUCH2 INTERES2 = Respons2
Let the errors of NOSAY1 and NOSAY2 correlate
Let the errors of COMPLEX1 and COMPLEX2 correlate
Let the errors of NOCARE1 and NOCARE2 correlate
Let the errors of TOUCH1 and TOUCH2 correlate
Let the errors of INTERES1 and INTERES2 correlate
Efficac2 = Efficac1
Respons2 = Respons1
Let the errors of Efficac2 and Respons2 correlate
LISREL Output: ND=3 SC MI2S ME=ULS IX=7427 NM=10
Path Diagram
End of Problem
```

- Line 1 specifies the raw data source.
- Lines 2 and 3 specify labels for the latent variables of the model.
- Lines 4 to 16 specify the two-wave model for the latent variables political efficacy and political responsiveness.

- Line 17 requests that the results in the output file should be given in terms of the LISREL model for the structural equation model (LISREL Output). It also requests that the results should be written to three decimal places (ND = 3), that the completely standardized solution should be written to the output file (SC), and robust unweighted least squares estimation (ME = ULS).
- Line 18 requests a path diagram of the model.
- Line 19 indicates that no more SIMPLIS commands are to be processed.

If the above **SPL** file is opened with LISREL and the **Run LISREL** button is clicked, the following path diagram is obtained.




The corresponding output file, **PANELUSA3A.OUT**, is opened in a separate window. A small portion of this file is shown in the following image.

```
PANELUSA3A.OUT
                              Goodness-of-Fit Statistics

 Degrees of Freedom for (C1)-(C3),C(5)-C(6)           24
 Maximum Likelihood Ratio Chi-Square (C1)             59.429 (P = 0.0001)
 Browne's (1984) ADF Chi-Square (C2_NT)               42.730 (P = 0.0107)
 Browne's (1984) ADF Chi-Square (C2_NNT)              18.672 (P = 0.7692)
 Satorra-Bentler (1988) Scaled Chi-Square (C3)        23.691 (P = 0.4794)
 Satorra-Bentler (1988) Adjusted Chi-Square           21.031 (P = 0.4760)
 Degrees of Freedom for C4                            21.305
 Chi-Square Scaled and Shifted (C5)                   23.709 (P = 0.4783)
 Yuan-Bentler (1997) Chi-Square for C2_NNT (C6)       18.306 (P = 0.7879)
 P-Value of C1 under Non-Normality                           = 0.4721
```

The Chi-square test statistic values for non-normality (C2_NNT, C3, C4, and C6) above indicate that the theoretical two-wave model for political efficacy and political responsiveness is supported by the data.

## 4. References


Aish, A.M. & Jöreskog, K.G. (1990). A panel model for political efficacy and responsiveness: An application of LISREL 7 with weighted least squares. *Quality and Quantity*, **24**, 405-426.

Browne, M.W. (1974). Generalised Least Squares Estimators in the analysis of covariance structures. *South African Statistical Journal*, **8**, 1-24. (Reprinted in Aigner, D.J. & Goldberger, A.S. (Eds.), *Latent variables in Socioeconomic Models*, pp. 205-226. Amsterdam: North Holland.)

Browne, M.W. (1984). Asymptotically Distribution-free Methods in the Analysis of Covariance Structures. *British Journal of Mathematical and Statistical Psychology*, **37**, 62-83.

Browne, M.W. (1987). Robustness of statistical inference in factor analysis and related models. Biometrika, 74, 375-384.

Chung, S, & Cai, L. (2019). Alternative Multiple Imputation Inference for Categorical Structural Equation Modeling. *Multivariate Behavioral Research*, **54(3)**, 323–337.

Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society, Series B*, **39**, 1–38.

Duncan, O.D., Haller, A.O. & Portes, A. (1971). Peer Influences on Aspirations, a Reinterpretation. *American Journal of Sociology*, **74**, 119-137.

Du Toit, M. & Du Toit, S.H.C. (2001). *Interactive LISREL: User's Guide*. Lincolnwood, IL: Scientific Software International, Inc.

Hasselrot, T. & Lernberg, L.O., (Eds.) (1980). *Tonåringen och livet*. Vällingby, Sweden: Liber förlag (in Swedish).

Jöreskog, K.G. (1981). Analysis of covariance structures. *Scandinavian Journal of Statistics*, **8**, 65-92.

Jöreskog, K. G. (1994). On the estimation of polychoric correlations and their asymptotic covariance matrix. *Psychometrika*, **59(3)**, 381–389.

Jöreskog, K.G. & Sörbom, D. (1999). *LISREL 8: User's Reference Guide*. Lincolnwood, IL: Scientific Software International, Inc.

Jöreskog, K.G., Sörbom, D., du Toit, S., & du Toit, M. (2001). LISREL 8: New statistical features. Lincolnwood, IL: Scientific Software International, Inc.

Mare, R.D. & Mason, W.M. (1981). Children's report of parental socioeconomic status. In G.W. Bohrnstedt & E.F. Borgatta (Eds.), *Social Measurement: Current Issues*. Beverly Hills: Sage Publications.

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, **44(4)**, 443–460.

Raudenbush, S., Bryk, A, Cheong, Y.F., Congdon, R & Du Toit (2011). HLM 7: Hierarchical Linear & Nonlinear Modeling. Lincolnwood, IL: Scientific Software International, Inc.

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. In *ASA 1988 Proceedings of the Business and Economic Statistics Section* (pp. 308–313). Alexandria, VA: American Statistical Association.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), Latent variables analysis: Applications for developmental research (pp. 399–419). Thousand Oaks, CA: Sage.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, **66**, 507–514.

Satorra, A., & Bentler, P. M. (2010). Ensuring positiveness of the scaled chi-square test statistic. *Psychometrika*, **75**, 243–248.

Steiger, J.H. & Lind, J.C. (1980). *Statistically-based tests for the number of common factors*. Paper presented at the annual meeting of the Psychometric Society, Iowa City, IA.