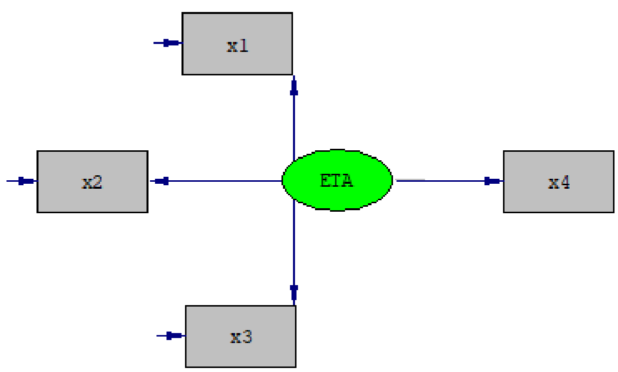
Several sets of congeneric measures: the multi-factor model

The most common type of measurement model is the one-factor congeneric measurement model, see Jöreskog (1971b). A path diagram of this model is shown in the figure below.



The corresponding equations are written in matrix form as

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \end{bmatrix} \xi + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \end{bmatrix}$$

or

$$\mathbf{x} = \boldsymbol{\lambda}\xi + \boldsymbol{\delta}.$$

The model is empirically not directly verifiable since there are more unobserved variables than observed. However, with the assumption that the latent variable is standardized, the equations imply that the covariance matrix of the observed variables is of the form

$$\mathbf{\Sigma} = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{\Theta} = \begin{bmatrix} \lambda_1^2 + \theta_{11} & & & \\ \lambda_2\lambda_1 & \lambda_2^2 + \theta_{22} & & \\ \lambda_3\lambda_1 & \lambda_3\lambda_2 & \lambda_3^2 + \theta_{33} & \\ \lambda_4\lambda_1 & \lambda_4\lambda_2 & \lambda_4\lambda_3 & \lambda_4^2 + \theta_{44} \end{bmatrix}$$

In this equation, $\mathbf{\Theta}$ is a diagonal matrix with elements θ_{ii}, the variances of δ_i (i = 1, 2, 3, 4).

The hypothesis that the population covariance matrix has this form is testable from a random sample of observations. In addition, the following subhypotheses are testable.

The above model is called the *congeneric measurement* model. The measures $x_1, x_2, ..., x_q$ are said to be *congeneric* if their true values $\tau_1, \tau_2, ..., \tau_q$ have all pair-wise correlations equal to unity. This is true of the model, since $\tau_i = \lambda_i \xi = x_i - \delta_i$ (i = 1, 2, 3, 4) and all τ's are linearly related and hence have unit correlation. The true variance in x_i is λ_i^2 and the reliability of x_i is

$$\rho_{ii} = \frac{\lambda_i^2}{\delta_{ii}} = \frac{\lambda_i^2}{\lambda_i^2 + \theta_{ii}} = 1 - \frac{\theta_{ii}}{\lambda_i^2 + \theta_{ii}}.$$

Strictly speaking, the error δ_i is considered to be sum of two uncorrelated random components s_i and e_i is the true measurement error. However, unless there are several replicate measures x_i with the same s_i, one cannot distinguish between these two components or separately estimate their variances. In consequence, ρ_{ii} is a lower bound for the true reliability.

Parallel measures have equal true score variances and equal error variances, i.e.,

$$\lambda_1^2 = \cdots = \lambda_4^2 \qquad \theta_{11} = \cdots = \theta_{44}.$$

Tau-equivalent measures have equal true score variances, but possibly different error variances.

Tests of parallelism and tau-equivalence are demonstrated in the following example.

The previous model generalizes immediately to several sets of congeneric measures. If there are n sets of such measures, with $m_1, m_2, ..., m_n$ in each set, respectively, we write

$$\mathbf{x}' = \left(x_1', x_2', ..., x_n'\right),$$

where $\mathbf{x}_g$ (g = 1, 2, …, n) is the vector of observed variables for the g-th set. Associated with the vector $\mathbf{x}_g$ there is a true score ξ_g and vectors $\boldsymbol{\lambda}_g$ and $\boldsymbol{\delta}_g$ defined as previously so that

$$\mathbf{x}_g = \boldsymbol{\lambda}_g \xi_g + \boldsymbol{\delta}_g .$$

As before we may assume, without loss of generality, that ξ_g is scaled to zero mean and unit variance. If the different latent variables $\xi_1, \xi_2, ..., \xi_n$ are all mutually uncorrelated, then each set of measures can be analyzed separately as in the previous section. However, in most cases these latent variables correlate with each other and an overall analysis of the entire set of measures must be made. Let

$$q = m_1 + m_2 + \cdots + m_n$$

be the total number of measurements. Then $\mathbf{x}$ is of order q. Let $\boldsymbol{\delta}$ be the corresponding vector of error scores. Furthermore, let

$$\xi' = \left[\xi_1, \xi_2, ..., \xi_n\right]$$

and let $\boldsymbol{\Lambda}$ be the matrix of order q x n, partitioned as

$$\boldsymbol{\Lambda} = \begin{bmatrix} \boldsymbol{\lambda}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\lambda}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \boldsymbol{\lambda}_n \end{bmatrix}.$$

Then $\mathbf{x}$ is represented as

$$\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta}.$$

Let $\boldsymbol{\Phi}$ be the correlation matrix of $\boldsymbol{\xi}$. Then the covariance matrix $\boldsymbol{\Sigma}$ of $\mathbf{x}$ is

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}\boldsymbol{\Phi}\boldsymbol{\Lambda}' + \boldsymbol{\Theta},$$

where $\boldsymbol{\Theta}$ is a diagonal matrix of order q containing the error variances.

In this example from Calsyn & Kenny (1977), the measured variables are

x_1 = self-concept of ability (S-C ABIL)

x_2 = perceived parental evaluation (PPAREVAL)

x_3 = perceived teacher evaluation (PTEAEVAL)

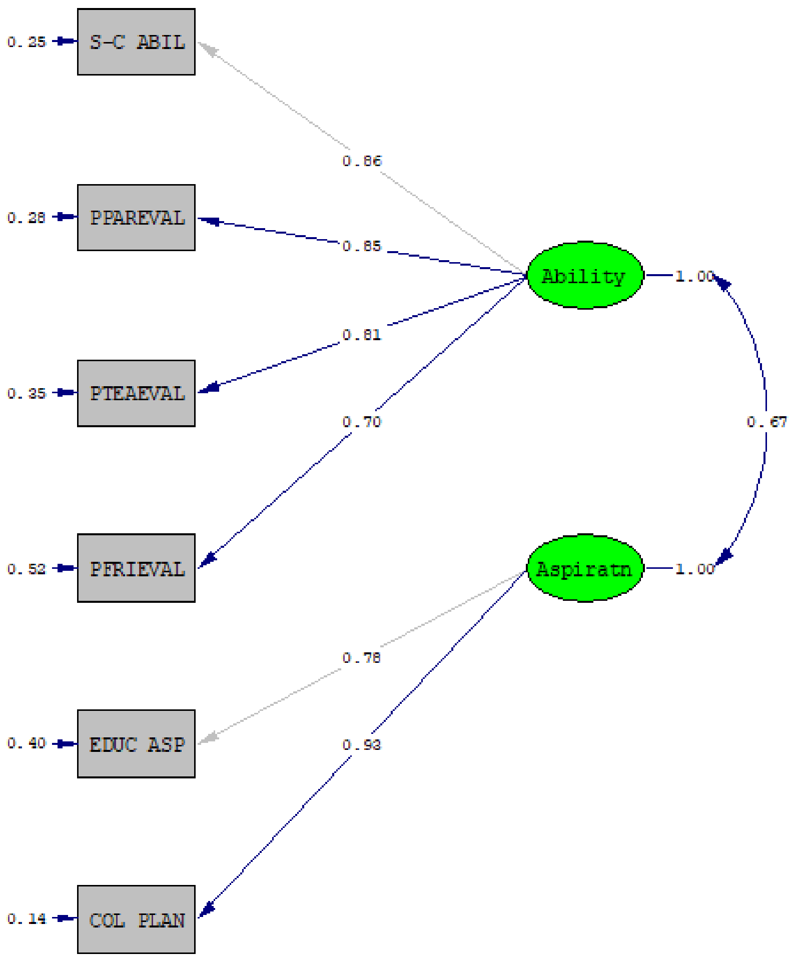
x_4 = perceived friend's evaluation (PFRIEVAL)

x_5 = educational aspiration (EDUC ASP)

x_6 = college plans (COL PLAN).

The x_1, x_2, x_3 and x_4 are assumed to be indicators of ability" and x_5 and x_6 are assumed to be indicators of "aspiration". The problem is to estimate the correlation between true ability and true aspiration.

The path diagram for this example is given below.



This model corresponds to the equations

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \end{bmatrix} = \begin{bmatrix} \lambda_{11} & 0 \\ \lambda_{21} & 0 \\ \lambda_{31} & 0 \\ \lambda_{41} & 0 \\ 0 & \lambda_{52} \\ 0 & \lambda_{62} \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} + \begin{bmatrix} \delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \end{bmatrix}$$

or

$$\mathbf{x} = \boldsymbol{\Lambda}\boldsymbol{\xi} + \boldsymbol{\delta}.$$

Note that the subscripts on λ correspond to the row and column of $\boldsymbol{\Lambda}$ where the coefficient appears.

The model includes only the three parameter matrices $\boldsymbol{\Lambda}$, $\boldsymbol{\Phi}$ and $\boldsymbol{\Theta}$, where $\boldsymbol{\Lambda}$ is a 6 x 2 matrix, $\boldsymbol{\Phi}$ is a 2 x 2 correlation matrix and $\boldsymbol{\Theta}$ is a 6 x 6 diagonal matrix with error variances in the diagonal. Consider first the solution in which the latent variables are in the metric of the reference variables. This is obtained by fixing a one in each column of $\boldsymbol{\Lambda}$.

Case 1:

For an analysis in which the latent variables are standardized, the LISREL command file (**EX32A.LIS** in the **LISREL Examples** folder) is as follows:

```
ABILITY AND ASPIRATION
DA NI=6 NO=556 MA=KM
LA
*
'S-C ABIL' PPAREVAL PTEAEVAL PFRIEVAL 'EDUC ASP' 'COL PLAN'
KM SY
(6F4.2)
 100
  73 100
  70  68 100
  58  61  57 100
  46  43  40  37 100
  56  52  48  41  72 100
MO NX=6 NK=2
LK
Ability Aspiratn
FR LX(2,1) LX(3,1) LX(4,1) LX(6,2)
VA 1 LX(1,1) LX(5,2)
OU SE TV RS MR FS
```

Note that labels containing blanks must be enclosed in single quotes.

The SY option of the KM command indicates that the lower half of the correlation matrix will be read with each row beginning a new record. The correlations are read in the fixed format following the KM command.

The MO command specifies six observed and two latent variables, but the model matrices are allowed to default as follows: all elements of $\mathbf{\Lambda}$ fixed initially at zero; $\mathbf{\Phi}$ is free symmetric; and $\mathbf{\Theta}$ is free diagonal.

The FR command frees four elements of $\mathbf{\Lambda}$.

The VA command fixes λ_{11} and λ_{52} at 1.0.

The OU command requests standard errors (SE), *t*-values (TV), fitted covariance (or correlation) matrix and fitted residuals (RS), covariances (or correlations) between observed and latent variables (MR), and factor scores regression (FS). The default maximum likelihood method of analysis (ML) is assumed.

Case 2:

For an analysis in which the two latent variables are standardized, the command file (**EX32B.LIS**) is:

```
ABILITY AND ASPIRATION
DA NI=6 NO=556 MA=KM
LA
*
'S-C ABIL' PPAREVAL PTEAEVAL PFRIEVAL 'EDUC ASP' 'COL PLAN'
KM SY
(6F4.2)
 100
  73 100
  70  68 100
  58  61  57 100
  46  43  40  37 100
  56  52  48  41  72 100
MO NX=6 NK=2
LK
ABILITY ASPIRATN
FR LX(1,1) LX(2,1) LX(3,1) LX(4,1) LX(5,2) LX(6,2)
OU SE TV RS MR FS
```

The input file differs from the previous in three ways: PH = SDT is added in MO, LX(1,1) and LX(5,2) have been added on the FR command, and the VA command has been deleted. PH = SDT fixes the scales for the latent variables so that $\mathbf{\Phi}$ becomes a correlation matrix, i.e., the two latent variables are standardized. The parameters λ_{11} and λ_{51} are now estimated instead of being fixed at 1. Note that the number of free parameters in the model remains the same.

The estimated $\mathbf{\Lambda}$ and $\mathbf{\Phi}$ for the two solutions are shown in the below. These solutions are equivalent. They only differ in the sense that the unit of measurement in ξ_1 and ξ_2 is different. The first model produces the output:

```
LISREL Estimates (Maximum Likelihood)

         LAMBDA-X

             Ability   Aspiratn
            --------   --------
 S-C ABIL      1.000       - -

 PPAREVAL      0.984       - -
             (0.040)
              24.312

 PTEAEVAL      0.933       - -
             (0.041)
              22.551

 PFRIEVAL      0.805       - -
             (0.044)
              18.289

 EDUC ASP       - -       1.000

 COL PLAN       - -       1.198
                        (0.073)
                         16.357


         PHI

             Ability   Aspiratn
            --------   --------
  Ability      0.745
             (0.061)
              12.292

 Aspiratn      0.446      0.601
             (0.044)    (0.063)
              10.069      9.612
```

while the output for the standardized solution is:

```
LISREL Estimates (Maximum Likelihood)

         LAMBDA-X

             ABILITY   ASPIRATN
            --------   --------
 S-C ABIL      0.863       - -
             (0.035)
              24.583
```

```
 PPAREVAL      0.849       - -
             (0.035)
              23.980

 PTEAEVAL      0.805       - -
             (0.036)
              22.135

 PFRIEVAL      0.695       - -
             (0.039)
              18.013

 EDUC ASP       - -        0.775
                         (0.040)
                          19.223

 COL PLAN       - -        0.929
                         (0.039)
                          23.592


         PHI

             ABILITY   ASPIRATN
            --------   --------
  ABILITY      1.000

 ASPIRATN      0.666      1.000
             (0.031)
              21.547
```

If the scales for ξ-variables are fixed by PH = ST, the solution obtained is not unique, since each column of $\mathbf{\Lambda}_x$ may be multiplied by minus one. This means that the fit function has several minima with the same fit. Therefore, when good reference variables are available, it is better to fix the scales by assigning fixed ones in each column of $\mathbf{\Lambda}_x$.

The results also include the correlations between the latent variables (factors) and the observed variables, shown below, and the factor scores regressions, shown thereafter (both for the standardized solution).

```
Covariances

         X - KSI

            S-C ABIL   PPAREVAL   PTEAEVAL   PFRIEVAL   EDUC ASP   COL PLAN
            --------   --------   --------   --------   --------   --------
  ABILITY      0.863      0.849      0.805      0.695      0.516      0.619
 ASPIRATN      0.575      0.566      0.536      0.463      0.775      0.929
```

In this case both $\mathbf{x}$ and ξ are standardized, and the entries above are correlations. In factor analysis terminology, they comprise the *factor structure*, as distinguished from the *factor pattern*, $\hat{\mathbf{\Lambda}}$, in the standardized solution of the results above. That is, the factor pattern is $\hat{\mathbf{\Lambda}}$ and the factor structure is $\hat{\mathbf{\Lambda}}\hat{\mathbf{\Phi}}$.

It should be pointed out that the elements of the factor pattern, i.e., the elements of $\hat{\mathbf{\Lambda}}$ are not in general correlations even if both $\mathbf{x}$ and ξ are standardized. The elements of $\hat{\mathbf{\Lambda}}$ are regression coefficients and, as such, they can exceed the absolute value one even though both $\mathbf{x}$ and ξ are standardized. The elements of the factor structure, i.e., the covariances between $\mathbf{x}$ and ξ, on the other hand, will of course be correlations if both $\mathbf{x}$ and ξ are standardized.

The estimated joint covariance matrix of $\mathbf{x}$ and ξ is

$$\begin{bmatrix} \hat{\mathbf{\Lambda}}\hat{\mathbf{\Phi}}\hat{\mathbf{\Lambda}}' + \hat{\mathbf{\Theta}} & \\ \hat{\mathbf{\Phi}}\hat{\mathbf{\Lambda}}' & \hat{\mathbf{\Phi}} \end{bmatrix}.$$

The upper left part of this matrix is the estimate $\hat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$ after the model has been fitted. This matrix may be saved in a file by putting SI = *filename* on the OU command.

```
Factor Scores Regressions

         KSI

            S-C ABIL   PPAREVAL   PTEAEVAL   PFRIEVAL   EDUC ASP   COL PLAN
            --------   --------   --------   --------   --------   --------
  ABILITY      0.341      0.307      0.230      0.135      0.024      0.085
 ASPIRATN      0.043      0.038      0.029      0.017      0.205      0.717
```

The coefficients above, representing the estimated bivariate regression of ξ_1 and ξ_2 on all the observed variables, are computed by the formula

$$\mathbf{A} = \hat{\mathbf{\Phi}}\hat{\mathbf{\Lambda}}'\hat{\mathbf{\Sigma}}^{-1} .$$

The matrix $\mathbf{A}$ may be saved in a file and used to compute estimated factor scores $\hat{\xi}_\alpha$ for any person with observed scores $\mathbf{x}_\alpha$, say, by the formula

$$\hat{\xi}_\alpha = \mathbf{A}\, x_\alpha .$$

When the LISREL model involves both ξ- and η-variables the factor scores regression will be computed by regressing all the ξ- and η-variables on all the observed variables.