Second order factor analysis

The equation

$$\mathbf{y} = \boldsymbol{\Lambda}_y \boldsymbol{\eta} + \boldsymbol{\varepsilon}$$

is in the form of a factor analysis model for $\mathbf{y}$ with first-order factors $\boldsymbol{\eta}$ and measurement errors $\boldsymbol{\varepsilon}$. Now suppose that the variables $\boldsymbol{\eta}$ in turn can be accounted for by a set of factors $\boldsymbol{\xi}$, so-called second-order factors, so that

$$\boldsymbol{\eta} = \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta},$$

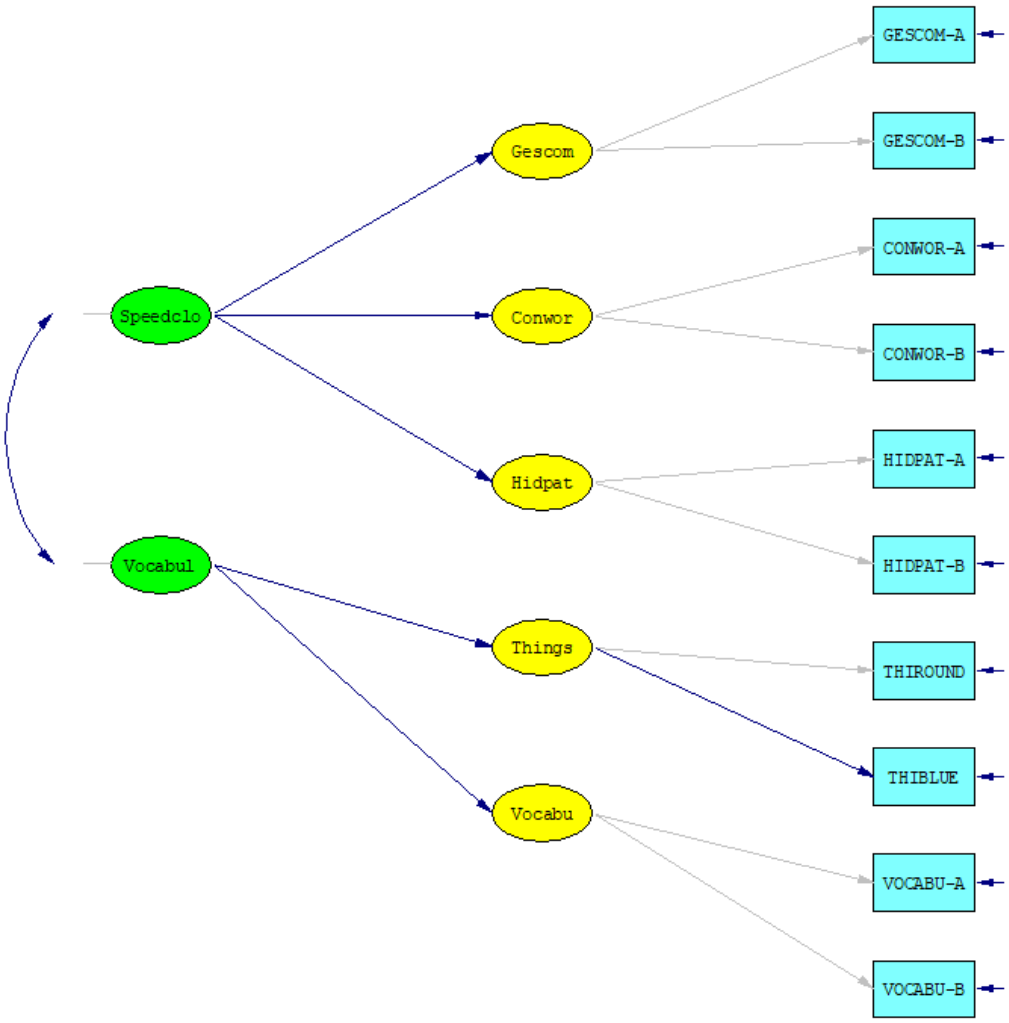
where $\boldsymbol{\Gamma}$ is a matrix of second-order factor loadings and $\boldsymbol{\zeta}$ is a vector of unique variances for $\boldsymbol{\eta}$. Combining these equations gives

$$\mathbf{y} = \boldsymbol{\Lambda}_y \left(\boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta} \right) + \boldsymbol{\varepsilon},$$

with covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda}_y \left(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi} \right) \boldsymbol{\Lambda}_y' + \boldsymbol{\Theta}_\varepsilon .$$

A path diagram for second-order factor analysis is shown below.



To illustrate the model, we use data on some cognitive ability tests. The standard deviations and correlations of two forms of each of five tests are given in the table following the path diagram. The sample size is 267.

Table: Correlations and standard deviations for some cognitive tests

Tests	Std. Dev.	Correlations									
GESCOM-A	2.42	1									
GESCOM-B	2.80	0.74	1								
CONWOR-A	3.40	0.33	0.42	1							
CONWOR-B	3.19	0.34	0.39	0.65	1						
HIDPAT-A	1.94	0.26	0.21	0.15	0.18	1					
HIDPAT-B	1.79	0.23	0.24	0.22	0.21	0.77	1				
THIROUND	5.63	0.15	0.12	0.14	0.11	0.17	0.20	1			
THIBLUE	3.10	0.14	0.14	0.14	0.15	0.06	0.09	0.42	1		
VOCABU-A	3.05	-0.04	-0.03	0.09	0.16	0.06	0.09	0.19	0.21	1	
VOCABU-B	2.25	0.02	0.02	0.10	0.23	0.04	0.07	0.09	0.21	0.72	1

We shall examine the hypothesis that the two forms of each test are tau-equivalent, except for the two-word fluency tests "Things Round" and "Things Blue" which are only assumed to be congeneric. The five true scores are postulated to depend on two factors, the first, "Speed of Closure", being measures by the first three tests and the second, "Vocabulary", being measured by the last two tests.

The model specification is:

$$\mathbf{\Lambda}_y = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & * & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{\Gamma} = \begin{bmatrix} * & 0 \\ * & 0 \\ * & 0 \\ 0 & * \\ 0 & * \end{bmatrix}, \quad \mathbf{\Phi} = \begin{bmatrix} 1 & \\ * & 1 \end{bmatrix},$$

$\mathbf{\Psi}$ is diagonal, and $\mathbf{\Theta}_\varepsilon$ is diagonal.

The LISREL command file (**EX62.LIS** in the **LISREL Examples** folder) is:

```
SECOND-ORDER FACTOR ANALYSIS
DA NI=10 NO=267
LA FI=EX62.DAT
KM FI=EX62.DAT
```

```
SD FI=EX62.DAT
MO NY=10 NE=5 NK=2 GA=FI PH=ST
LE
Gescom Conwor Hidpat Things Vocabu
LK
Speedclo Vocabul
VA 1 LY 1 1 LY 2 1 LY 3 2 LY 4 2 LY 5 3 LY 6 3 LY 7 4 LY 9 5 LY 10 5
FR LY 8 4 GA 1 1 GA 2 1 GA 3 1 GA 4 2 GA 5 2
ST 1 ALL
OU SE TV SS NS
```

The χ^2 goodness-of-fit statistic obtained is shown below.

```
                    Goodness-of-Fit Statistics

 Degrees of Freedom for (C1)-(C2)                    33
 Maximum Likelihood Ratio Chi-Square (C1)            53.255 (P = 0.0142)
```

This does not represent a particularly good fit. The lack of fit may be due to either the assumption of tau-equivalence for the four pairs of test forms or the hypothesized second-order structure. When the assumption of tau-equivalence is relaxed, χ^2 drops to 41.69 with 29 degrees of freedom. The difference in these, with 4 degrees of freedom, has a probability level of 0.025. This suggests that the lack of fit is probably due to both assumptions.

By computing the diagonal elements of $\hat{\boldsymbol{\Lambda}}_y \hat{\boldsymbol{\Gamma}} \hat{\boldsymbol{\Phi}} \hat{\boldsymbol{\Gamma}}' \hat{\boldsymbol{\Lambda}}_y'$, $\hat{\boldsymbol{\Lambda}}_y \hat{\boldsymbol{\Psi}} \hat{\boldsymbol{\Lambda}}_y'$, and $\hat{\boldsymbol{\Theta}}_\varepsilon$ and relating these to the total variance in $\hat{\boldsymbol{\Sigma}}$, one gets a variance decomposition of each test form into common, specific and error components, respectively. These variance components are shown in the table below.

Table: Various test theory models

I	Error	Specific	Common
GESCOM-A	0.193	0.428	0.379
GESCOM-B	0.338	0.351	0.311
CONWOR-A	0.386	0.279	0.335
CONWOR-B	0.313	0.313	0.374
HIDPAT-A	0.291	0.587	0.122
HIDPAT-B	0.163	0.694	0.144
THIROUND	0.682	0.145	0.172
THIBLUE	0.445	0.254	0.301
VOCABU-A	0.468	0.465	0.067
VOCABU-B	0.044	0.835	0.121