*Kimmo Vehkalahti* (4 April 2013):

# The matrices of SEM: a few words (and formulas)

The following is primarily targeted for the students of Statistics, Economics, Mathematics etc., who are familiar with the matrix algebra and notation. For the students of other disciplines, this may well be considered as "nice to know".

In any case, matrices are the key for understanding the Structural Equation Models (or any other statistical models) on more advanced level. Here, I use the traditional (or "LISREL") matrix notation, which was made famous by *Karl Jöreskog* in the 1970s. The focus here is on the covariance structures, and hence no means (latent or observed) appear in the equations. Adding the mean terms would require introducing just a few more Greek letters...

## Three fundamental equations

In general, **three equations** are needed. They will include altogether **seven vectors** and **eight matrices**. Let us have a look at all of them.

The first equation gives the **structural model**

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\boldsymbol{\xi} + \boldsymbol{\zeta}, \tag{1}$$

where $\boldsymbol{\eta}$ *(eta)* is an $m \times 1$ vector of the latent dependent variables and $\boldsymbol{\xi}$ *(ksi)* is an $n \times 1$ vector of the latent independent variables.

The relationships of the latent variables are denoted by the matrices $\boldsymbol{B}$ *(capital beta)* and $\boldsymbol{\Gamma}$ *(capital gamma)* so that $\boldsymbol{B}$ is an $m \times m$ matrix of structure coefficients relating the latent dependent variables ($\boldsymbol{\eta}$) to one another, and $\boldsymbol{\Gamma}$ is an $m \times n$ matrix of structure coefficients that relate the latent independent variables ($\boldsymbol{\xi}$) to the latent dependent variables ($\boldsymbol{\eta}$). The $m \times 1$ vector $\boldsymbol{\zeta}$ *(zeta)* represents the prediction errors.

The $n \times n$ covariance matrix of the latent independent variables ($\boldsymbol{\xi}$) is denoted by $\boldsymbol{\Phi}$ *(capital phi)*. The $m \times m$ covariance matrix of the prediction errors ($\boldsymbol{\zeta}$) is denoted by $\boldsymbol{\Psi}$ *(capital psi)*.

The other two equations are needed for the **measurement models:**

$$\boldsymbol{y} = \boldsymbol{\Lambda}_y\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{2}$$

for the latent dependent variables $\boldsymbol{\eta}$, and

$$\boldsymbol{x} = \boldsymbol{\Lambda}_x\boldsymbol{\xi} + \boldsymbol{\delta}, \tag{3}$$

for the latent independent variables $\boldsymbol{\xi}$. The observed items are denoted by the $p \times 1$ vector $\boldsymbol{y}$ for the measures of the latent dependent variables ($\boldsymbol{\eta}$), and by the $q \times 1$ vector $\boldsymbol{x}$ for the measures of the latent independent variables ($\boldsymbol{\xi}$).

The relationships between the items and the latent variables (in other words, the factor loadings) are denoted by the $p \times m$ matrix $\boldsymbol{\Lambda}_y$ *(capital lambda sub y)* and the $q \times n$ matrix $\boldsymbol{\Lambda}_x$ *(capital lambda sub x)*.

The measurement errors for the items $\boldsymbol{y}$ are denoted by the $p \times 1$ vector $\boldsymbol{\varepsilon}$ *(epsilon)* and for the items $\boldsymbol{x}$ by the $q \times 1$ vector $\boldsymbol{\delta}$ *(delta)*. The covariance matrices of these errors are denoted by $\boldsymbol{\Theta}_\varepsilon$ *(capital theta sub epsilon)*, which is a $p \times p$ matrix, and $\boldsymbol{\Theta}_\delta$ *(capital theta sub delta)*, which is a $q \times q$ matrix.

## Vectors and matrices

The seven vectors and eight matrices we have defined above are:

- $\boldsymbol{\eta}$ $(m \times 1)$
- $\boldsymbol{\xi}$ $(n \times 1)$
- $\boldsymbol{\zeta}$ $(m \times 1)$
- $\boldsymbol{y}$ $(p \times 1)$
- $\boldsymbol{\varepsilon}$ $(p \times 1)$
- $\boldsymbol{x}$ $(q \times 1)$
- $\boldsymbol{\delta}$ $(q \times 1)$

and

- $\boldsymbol{B}$ $(m \times m)$
- $\boldsymbol{\Gamma}$ $(m \times n)$
- $\boldsymbol{\Phi}$ $(n \times n)$
- $\boldsymbol{\Psi}$ $(m \times m)$
- $\boldsymbol{\Lambda}_y$ $(p \times m)$
- $\boldsymbol{\Lambda}_x$ $(q \times n)$
- $\boldsymbol{\Theta}_\varepsilon$ $(p \times p)$
- $\boldsymbol{\Theta}_\delta$ $(q \times q)$

## The covariance matrix $\boldsymbol{\Sigma}$

The covariance matrix $\boldsymbol{\Sigma}$ implied by the model may be denoted by a supermatrix

$$\begin{bmatrix} \boldsymbol{\Sigma}_{yy} & \boldsymbol{\Sigma}_{yx} \\ \boldsymbol{\Sigma}_{xy} & \boldsymbol{\Sigma}_{xx} \end{bmatrix}, \tag{4}$$

where the four submatrices are defined as follows.

The matrix $\boldsymbol{\Sigma}_{yy}$ in the upper left corner deals with the covariance terms within the $\boldsymbol{y}$ variables. It can be written in the form

$$\boldsymbol{\Sigma}_{yy} = \boldsymbol{\Lambda}_y(\boldsymbol{I} - \boldsymbol{B})^{-1}(\boldsymbol{\Gamma}\boldsymbol{\Phi}\boldsymbol{\Gamma}' + \boldsymbol{\Psi})(\boldsymbol{I} - \boldsymbol{B}')^{-1}\boldsymbol{\Lambda}'_y + \boldsymbol{\Theta}_\varepsilon, \tag{5}$$

where $\boldsymbol{I}$ is an $m \times m$ identity matrix.

Correspondingly, the matrix $\boldsymbol{\Sigma}_{xx}$ in the lower right corner deals with the covariance terms within the $\boldsymbol{x}$ variables. It can be written in the form

$$\boldsymbol{\Sigma}_{xx} = \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Lambda}'_x + \boldsymbol{\Theta}_\delta, \tag{6}$$

which is, by the way, the usual equation of the common factor analysis model.

Finally, the matrix $\boldsymbol{\Sigma}_{xy}$ in the lower left corner (or its transpose $\boldsymbol{\Sigma}_{yx}$ in the upper right corner) deals with the covariance terms between the $\boldsymbol{x}$ and $\boldsymbol{y}$ variables. It can be written in the form

$$\boldsymbol{\Sigma}_{xy} = \boldsymbol{\Lambda}_x\boldsymbol{\Phi}\boldsymbol{\Gamma}'(\boldsymbol{I} - \boldsymbol{B}')^{-1}\boldsymbol{\Lambda}'_y. \tag{7}$$