Psicológica (2004), 25, 253-274

# Structural Equation Modelling of Multiple Facet Data: Extending Models for Multitrait-Multimethod Data

Timo M. Bechger and Gunter Maris
CITO (The Netherlands)

**Abstract**

This paper is about the structural equation modelling of quantitative measures that are obtained from a multiple facet design. A facet is simply a set consisting of a finite number of elements. It is assumed that measures are obtained by combining each element of each facet. Methods and traits are two such facets, and a multitrait-multimethod study is a two-facet design. We extend models that were proposed for multitrait-multimethod data by Wothke (1984;1996) and Browne (1984, 1989, 1993), and demonstrate how they can be fitted using standard software for structural equation modelling. Each model is derived from the model for individual measurements in order to clarify the first principles underlying each model.

## Introduction

A *Multi-Trait Multi-Method (MTMM)* study is characterized by measures that are composed as combinations of traits and methods. In this paper, we will treat a more general case where measures are composed as combinations of elements of facets. A *facet* is simply a set consisting of a finite

Address correspondence to: Timo Bechger, CITO, P.O. Box 1034, NL-6801 MG, Arnhem, The Netherlands. E-mail: timo.bechger@citogroep.nl; Tel:+31-026-3521162.

number of elements, usually called *conditions*. Facets refer to properties of the measures or measurement conditions. Methods and traits are two such facets, and a MTMM study is a (fully crossed) two-facet design. Facets need not be methods or traits or anything in particular. Consider, for example, a study that is presented by Browne (1970) and discussed in detail by Jöreskog and Sörbom (1996, section 6.3). Persons were seated in a darkened room and required to place a rod in vertical position by pushing buttons. The score was the (positive or negative) angle of the rod from the vertical. Each person had to perform the task twice in a number of different situations which where constructed according to a two-facet design. The two facets were the position of the chair and the initial position of the rod, each with three conditions. The occasion of the experiment may be considered the third facet with two conditions.

We assume that measures were constructed for each combination of the facets and that we have data for each measure. We further assume that measures are continuous. We believe that the case of discrete data is more appropriately handled using item response theory models (e.g., Bechger, Verhelst & Verstralen, 2001).

To analyze data from a multiple facet design, we extend two models that were suggested for the MTMM design: the *covariance component model* (Wothke, 1984, 1996), and the *composite direct product model* (Browne, 1984, 1989, 1993). In doing so, we pursue e.g., Bagozzi, Yi and Nassen, (1999), Cudeck (1988), or Browne and Strydom (1997) who suggest generalization of the composite direct product model to multiple facets. Our objective is to demonstrate how researchers who know the basic principles of *structural equation modelling* (SEM) may formulate and fit these models using the LISREL (Jöreskog & Sörbom, 1996) or the Mx program[1] (Neale, Boker, Xie, & Maes, 2002). There are several alternative software packages but the majority of these have an interface that is similar to that of LISREL or Mx. For a general introduction to SEM we refer the reader to Bollen (1989).

[1]The Mx program is free-ware and can, at present, be obtained from the internet address http://www.vcu.edu/mx/index.html

Each model is derived from the (data) model for individual measurements in order to clarify the first principles underlying each model. In the context of MTMM studies, the model for the observations is of less interest since the main objective is to establish a structure for the correlations that relates to the Campbell and Fiske (1959) criteria for convergent and discriminant validity. However, in general multi-facet studies the data model is important as a substantive hypothesis that guides the interpretation of the parameters. We demonstrate how each of the models can be fitted to a correlation or covariance matrix. (We refer to Cudeck (1989) for a survey of the issues concerning the analysis of correlation matrices). As an illustration, we discuss a number of applications to real data.

## Preliminaries

Persons (or, more general, *objects*) are assumed to be drawn *at random* from a large population and each observation is taken to be a realization of a random vector $\mathbf{x}$ of measurements made under combinations of conditions of multiple facets. All models that are considered here are based upon the following linear model for the observations

$$\mathbf{x} = \boldsymbol{\mu} + \boldsymbol{\eta} + \mathbf{u} \quad ,$$

where $\boldsymbol{\mu}$ denote the mean of $\mathbf{x}$, and the latent variable $\boldsymbol{\eta}$ represents true or common scores. The components of $\mathbf{u}$ represent measurement error and are assumed to be uncorrelated with mean zero and variance matrix $\mathbf{D}_u$. The common scores are uncorrelated to $\mathbf{u}$; they have zero mean and covariance matrix $\boldsymbol{\Sigma}_\eta$. It follows that the data have mean vector $\boldsymbol{\mu}$ and covariance matrix

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_\eta + \mathbf{D}_u \quad ,$$

where $\mathbf{D}_u$ is a diagonal matrix.

We use the following notation: In a design with multiple facets, A, B, C, D, etc. denote the facets. Each facet has several conditions (or elements) denoted by $\mathrm{A}_i$, $\mathrm{B}_j$, etc. The number of conditions in each facet is denoted by the lowercase of the letter that is used to denote the facet. For example, $a = 3$ if facet A has three conditions. We use $F$ to denote a generic facet. The number of measures that can be constructed (e.g., $a \times b \times c$ if there are three facets) will be denoted by $p$. The number of facets will be denoted by #F. The sum of the conditions in each of the facets (e.g., $a + b + c$) will be denoted by #f. The symbol $\mathbf{I}_a$ denotes an identity matrix of dimension $a$ and $\mathbf{1}_a$ a unit vector with $a$ elements. The symbol $\otimes$ denotes the Kronecker or direct product operator with $\mathbf{A}\otimes\mathbf{B} = (a_{ij}\mathbf{B})$. Finally, the vector $\mathbf{z}_F$ denotes a vector of random variables associated with a facet; that is, $\mathbf{z}_F = (z(F_1), \ldots, z(F_f))^T$, where $z(F_j)$ denotes a random variable associated with the j-th condition of facet F. Uppercase $T$ refers to transposition.

## The Covariance Component Model

Wothke (1984) suggested that the *Covariance Component* (CC) model described by Bock and Bargmann (1966) be used for MTMM data. In this section, we discuss a number of parameterizations of the CC model and demonstrate how each is specified within the LISREL framework. Note that the CC model is related to random effects analysis of variance (see Bock & Bargmann, 1966, pp. 508-509) but we will not explicitly use this relationship in our presentation of the model.

### *Introduction*

Let $\eta_x$ denote a generic element of $\boldsymbol{\eta}$; i.e., x is a measurement obtained as the combination of $A_i, B_j, \ldots, E_r$. In the CC model, $\boldsymbol{\eta}$ is assumed to have an additive structure. Specifically,

$$\eta_x = g + z(A_i) + z(B_j) + \ldots + z(E_r) \quad ,$$

where $g$ denotes a within-person mean. In matrix notation:

$$\boldsymbol{\eta} = \mathbf{A}\mathbf{z} \quad ,$$

where $\mathbf{z} = \left(g, \mathbf{z}_A^T, \mathbf{z}_B^T, \ldots, \mathbf{z}_E^T\right)^T$, and $\mathbf{A}$ is a $p \times (1 + \#f)$ incidence matrix; that is, a matrix whose entries are zero or one. The rows of $\mathbf{A}$ indicate all combinations of conditions of each of the facets. For two to five facets, the structure of $\mathbf{A}$ is given in Equation 1.

$$\begin{array}{llllll}
[\mathbf{1}_p & \mathbf{I}_a \otimes \mathbf{1}_b & \mathbf{1}_a \otimes \mathbf{I}_b] & & & \\
[\mathbf{1}_p & \mathbf{I}_a \otimes \mathbf{1}_{bc} & \mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_c & \mathbf{1}_{ab} \otimes \mathbf{I}_c] & & \\
[\mathbf{1}_p & \mathbf{I}_a \otimes \mathbf{1}_{bcd} & \mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_{cd} & \mathbf{1}_{ab} \otimes \mathbf{I}_c \otimes \mathbf{1}_d & \mathbf{1}_{abc} \otimes \mathbf{I}_c] & \\
[\mathbf{1}_p & \mathbf{I}_a \otimes \mathbf{1}_{bcde} & \mathbf{1}_a \otimes \mathbf{I}_b \otimes \mathbf{1}_{cde} & \mathbf{1}_{ab} \otimes \mathbf{I}_c \otimes \mathbf{1}_{de} & \mathbf{1}_{abc} \otimes \mathbf{I}_d \otimes \mathbf{1}_e & \mathbf{1}_{abcd} \otimes \mathbf{I}_e]
\end{array} \tag{1}$$

For example, if there are two facets, with two conditions each:

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} \quad . \tag{2}$$

Note that Equation 1 was derived assuming that each subsequent facet is nested in the preceding facet(s); e.g., $A_1B_1C_1$, $A_1B_1C_2$, $A_1B_2C_1$, etc. It is easy to see the general pattern in Equation 1 and derive expressions for more than five facets.

It is assumed that $\mathbf{z}$ is multivariate normally distributed with zero mean and covariance matrix $\mathbf{\Sigma}_z$. Furthermore, each facet is assumed to have an independent influence on the measurements so that $\mathbf{\Sigma}_z$ is block-diagonal; that is,

$$\mathbf{\Sigma}_z = diag\left(\sigma_g^2, \mathbf{\Sigma}_A, \cdots, \mathbf{\Sigma}_E\right) = \begin{bmatrix} \sigma_g^2 & & & \cdots & \\ & \mathbf{\Sigma}_A & & \cdots & \\ & & \mathbf{\Sigma}_B & \cdots & \\ \vdots & \vdots & \vdots & \ddots & \\ & & & \cdots & \mathbf{\Sigma}_E \end{bmatrix} \quad , \tag{3}$$

where $\mathbf{\Sigma}_F$ denotes a within-facet dispersion matrix. It follows that

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Sigma}_z\mathbf{A}^T + \mathbf{D}_u \quad . \tag{4}$$

This is a *confirmatory factor analysis (CFA)* model with a constant factor loading matrix (see Bollen, 1989, chapter 7). As it stands, the model is not identifiable. Informally, this means that no amount of data will help to determine the true value of one or more of the parameters. We will demonstrate this by constructing an equivalent model with less parameters.

The matrix $\mathbf{A}$ has $1+\#f$ columns and rank equal to $r(\mathbf{A}) = 1+\#f-\#F$ (e.g., Equation 2). Since $\mathbf{A}$ has deficient column rank, the vector of random components that satisfies $\boldsymbol{\eta} = \mathbf{Az}$ need not be unique. Consider an example with two facets with two conditions each. Then, if $\mathbf{z}_1 = (1,2,3,4,5)^T$ and $\mathbf{z}_2 = (7,0,1,0,1)^T$, $\boldsymbol{\eta} = \mathbf{Az}_1 = \mathbf{Az}_2$.

If $\mathbf{z}$ is a solution to $\boldsymbol{\eta} = \mathbf{Az}$ we may write any other solution $\mathbf{z}^*$ as

$$\mathbf{z}^* = \begin{bmatrix} g + z(A_1) + z(B_1) \\ 0 \\ z(A_2) - z(A_1) \\ 0 \\ z(B_2) - z(B_1) \end{bmatrix} + \begin{bmatrix} 2v_1 + v_2 + v_4 \\ v_2 \\ 2v_3 - v_2 \\ v_4 \\ 2v_5 - v_4 \end{bmatrix} \tag{5}$$

where $v_1$ to $v_5$ are arbitrary constants (e.g., Pringle & Rainer, 1971, p. 10). The second and fourth elements of $\mathbf{z}^*$ are arbitrary which means that the corresponding entries of $\boldsymbol{\Sigma}_z$ are arbitrary and therefore not identifiable. Specifically, if $\boldsymbol{\Sigma}_z*$ denotes the covariance matrix of $\mathbf{z}^*$, it is easily checked that $\mathbf{A}\boldsymbol{\Sigma}_z*\mathbf{A}^T$ equals $\mathbf{A}\boldsymbol{\Sigma}_z\mathbf{A}^T$, where $\boldsymbol{\Sigma}_z$ has seven parameters and $\boldsymbol{\Sigma}_z*$ five (see Equation 8).

The first vector in (5) contains linear combinations $\boldsymbol{\xi} = \mathbf{Lz}$ of the random components that are common to all solutions. In general, $\mathbf{L}$ denotes a $r(\mathbf{A}) \times (1+\#f)$ matrix of full row rank. In the example[2],

$$\boldsymbol{\xi} = \begin{bmatrix} g + z(A_1) + z(B_1) \\ z(A_2) - z(A_1) \\ z(B_2) - z(B_1) \end{bmatrix} = \mathbf{Lz} = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \mathbf{z} \quad .$$

We will not give a general expression for $\mathbf{L}$ but note that, in general, the first linear combination $\xi_1$ is the common score of the first measurement. The other linear combinations are within-facet deviations from the first condition.

[2] It is easily checked that $\mathbf{Lz}_1 = \mathbf{Lz}_2 = (7,1,1)$.

We now wish to find an equivalent expression of $\boldsymbol{\eta}$ in terms of $\boldsymbol{\xi}$. That is, we look for a matrix $\boldsymbol{\Lambda}$ such that $\mathbf{Az} = \boldsymbol{\Lambda\xi}$. In our example,

$$\begin{aligned}
\mathbf{Az} &= \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} g \\ z(A_1) \\ z(A_2) \\ z(B_1) \\ z(B_2) \end{bmatrix} \\
&= \begin{bmatrix} g + z(A_1) + z(B_1) \\ g + z(A_1) + z(B_2) \\ g + z(A_2) + z(B_1) \\ g + z(A_2) + z(B_2) \end{bmatrix} \\
&= \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} g + z(A_1) + z(B_1) \\ z(A_2) - z(A_1) \\ z(B_2) - z(B_1) \end{bmatrix} = \boldsymbol{\Lambda\xi} \quad .
\end{aligned}$$

It is seen that $\boldsymbol{\Lambda}$ is equal to $\mathbf{A}$ with columns corresponding to the first condition in each facet deleted. In general, $\boldsymbol{\Lambda}$ has the same structure as $\mathbf{A}$ in Equation 1 except that the first column in each of the identity matrices in (1) is deleted; that is, $\mathbf{I}_f$ is replaced by $\begin{bmatrix} \mathbf{0}_{f-1} & \mathbf{I}_{f-1} \end{bmatrix}$.

It follows that

$$\boldsymbol{\Sigma} = \boldsymbol{\Lambda\Sigma}_\xi\boldsymbol{\Lambda}^T + \mathbf{D}_u \quad , \tag{6}$$

where $\boldsymbol{\Sigma}_\xi = \mathbf{L}\boldsymbol{\Sigma}_z\mathbf{L}^T$ denotes the dispersion matrix of $\boldsymbol{\xi}$. This CFA model is easily fitted with LISREL or Mx. Note that $\boldsymbol{\Sigma}_\xi$ does not inherit the block-diagonal structure of $\boldsymbol{\Sigma}_z$ because $\xi_1 = g + z(A_1) + \cdots + z(E_1)$ correlates to each of the within-facet deviations. However, if $\boldsymbol{\Sigma}_z$ has the postulated block-diagonal structure, the within-facet deviations should be uncorrelated between facets.

*Alternative Parameterizations I*

The matrix $\boldsymbol{\Lambda}$ was constructed by deleting columns from $\mathbf{A}$. Hence, $\boldsymbol{\Lambda}$ may be written as the matrix product $\mathbf{AP}$, where $\mathbf{P}$ is an incidence matrix that serves to delete columns from $\mathbf{A}$. It follows that we may express a model that is equivalent to (6) as:

$$\mathbf{\Sigma} = \mathbf{A}\mathbf{\Sigma}_{z^*}\mathbf{A}^T + \mathbf{D}_u \quad , \tag{7}$$

where $\mathbf{\Sigma}_{z^*} = \mathbf{P}\mathbf{\Sigma}_\xi\mathbf{P}^T$. In the example,

$$\mathbf{\Sigma}_{z^*} = \begin{bmatrix} E\left[\xi_1^2\right] & 0 & E\left[\xi_2,\xi_1\right] & 0 & E\left[\xi_3,\xi_1\right] \\ 0 & 0 & 0 & 0 & 0 \\ E\left[\xi_2,\xi_1\right] & 0 & E\left[\xi_2^2\right] & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ E\left[\xi_3,\xi_1\right] & 0 & 0 & 0 & E\left[\xi_3^2\right] \end{bmatrix} \quad . \tag{8}$$

Instead of fitting the model (6), we may fit a model of the form (7) with appropriate restrictions on $\mathbf{\Sigma}_{z^*}$. In general, these restrictions are that all entries in $\mathbf{\Sigma}_{z^*}$ involving the first condition of any of the facets are fixed to zero, as in (8).

*Alternative Parameterizations II*

Let $\mathbf{T}$ denote an arbitrary, non-singular matrix and define $\mathbf{\Lambda}^* = \mathbf{\Lambda}\mathbf{T}^{-1}$, and $\mathbf{\Sigma}_{\xi^*} = \mathbf{T}\mathbf{\Sigma}_\xi\mathbf{T}^T$. It is not difficult to see that any model that can be written as:

$$\mathbf{\Sigma} = \mathbf{\Lambda}^*\mathbf{\Sigma}_{\xi^*}\mathbf{\Lambda}^{*T} + \mathbf{D}_u \tag{9}$$

is equivalent to (6). The matrix $\mathbf{\Sigma}_{\xi^*}$ denotes the dispersion matrix of

$$\boldsymbol{\xi}^* = \mathbf{T}\boldsymbol{\xi} = \mathbf{T}\mathbf{L}\mathbf{z} = \mathbf{L}^*\mathbf{z} \quad .$$

Hence, each alternative model of the form (9) implies a set of linear combinations defined by $\mathbf{L}^* = \mathbf{T}\mathbf{L}$.

Browne (1989) considers linear combinations of the form:

$$
\begin{aligned}
\boldsymbol{\xi}^* &= \mathbf{L}^*\mathbf{z} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} \begin{bmatrix} g \\ z(A_1) \\ z(A_2) \\ z(B_1) \\ z(B_2) \end{bmatrix} \\
&= \begin{bmatrix} g + \frac{1}{2}\left(z(A_1) + z(A_2)\right) + \frac{1}{2}\left(z(B_1) + z(B_2)\right) \\ \frac{1}{2}\left(z(A_2) - z(A_1)\right) \\ \frac{1}{2}\left(z(B_2) - z(B_1)\right) \end{bmatrix} \quad .
\end{aligned}
$$

The corresponding matrix $\mathbf{T}$, is found be solving $\mathbf{L}^* = \mathbf{TL}$. Here,

$$
\mathbf{L}^* = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} & \frac{1}{2} \\ 0 & -\frac{1}{2} & \frac{1}{2} & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \mathbf{TL} = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix} \quad .
$$

The corresponding $\boldsymbol{\Lambda}^*$ is

$$
\boldsymbol{\Lambda}^* = \begin{bmatrix} 1 & -1 & -1 \\ 1 & -1 & 1 \\ 1 & 1 & -1 \\ 1 & 1 & 1 \end{bmatrix} = \boldsymbol{\Lambda}\mathbf{T}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & -1 & -1 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix} \quad .
$$

For later reference, we call this parametrization *Browne's parametrization*. In general, if we use Browne's parametrization, $\boldsymbol{\Lambda}^*$ has the same structure as $\mathbf{A}$ in Equation 1, except that each of the identity matrices $\mathbf{I}_f$ in (1) is replaced by the matrix $\begin{bmatrix} -\mathbf{1}_{f-1} & \mathbf{I}_{f-1} \end{bmatrix}^T$. There is, of course, an infinite number of alternative parameterizations, each corresponding to a non-singular matrix $\mathbf{T}$ and a matrix $\mathbf{L}^*$ (see e.g., Bock & Bargmann, 1966, table 6).

Browne (1989) implicitly assumes that sub-matrices $\boldsymbol{\Sigma}_F$ in (3) have equal diagonal elements. This implies that $\boldsymbol{\Sigma}_{\xi^*} = \mathbf{L}^*\boldsymbol{\Sigma}_z\mathbf{L}^{*T}$ is block-diagonal. For instance,

$$
\begin{aligned}
E\left[\xi_1^*, \xi_2^*\right] &= E\left[\frac{1}{2}\left(z(A_1) + z(A_2)\right), \frac{1}{2}\left(z(A_2) - z(A_1)\right)\right] \\
&= \frac{1}{4}\left(\sigma^2_{z(A_2)} - \sigma^2_{z(A_1)}\right) \quad ,
\end{aligned}
$$

which is zero if and only if $\sigma^2_{z(A_1)} = \sigma^2_{z(A_2)}$.

*Fitting the Covariance Component model to a Correlation Matrix*

To fit the model to a correlation matrix we need to derive the model for the correlation matrix $\mathbf{P} = \mathbf{D}_x \boldsymbol{\Sigma} \mathbf{D}_x$, where $\mathbf{D}_x = diag^{-\frac{1}{2}}(\boldsymbol{\Sigma})$; a diagonal matrix with on the diagonal the inverses of the population standard-deviations of the observed measures. That is,

$$\mathbf{P} = \mathbf{D}_x \left(\boldsymbol{\Sigma}_\eta + \mathbf{D}_u\right) \mathbf{D}_x = \mathbf{D}_x \left(\boldsymbol{\Lambda} \boldsymbol{\Sigma}_\xi \boldsymbol{\Lambda}^T + \mathbf{D}_u\right) \mathbf{D}_x \quad . \tag{10}$$

This model is easily fitted with LISREL. One may, as in (10), use our first parametrization and specify: ny = ne = p, nk= $r(\mathbf{A})$, PSI= $\mathbf{0}$, LAMBDA-Y $= \mathbf{D}_x$, GAMMA $= \boldsymbol{\Lambda}$, and PHI $= \boldsymbol{\Sigma}_\xi$. An example of a LISREL script is provided in the Appendix. Only small changes are necessary to specify $\boldsymbol{\Sigma}_\eta$ as in (7) or (9). Note that (10) is a special case of the scale-free covariance model proposed by Wiley, Schmidt, and Bramble (1973).

There is a caveat however. Wothke (1988; 1996) claims that the variance of $\xi_1$ and covariances involving $\xi_1$ are not identifiable. Hence, in general, model (10) is unsuited for correlation matrices. As mentioned before, Browne (1989) assumes that the within-facet dispersion matrices have equal diagonal elements. Suppose that this assumption holds. Then, if Browne's parametrization is used, and the matrix $\boldsymbol{\Sigma}_\eta$ in (10) is specified as $\boldsymbol{\Lambda}^* \boldsymbol{\Sigma}_{\xi^*} \boldsymbol{\Lambda}^{*T}$, the matrix $\boldsymbol{\Sigma}_{\xi^*}$ is block-diagonal and the model may be used for correlation matrices provided the variance of $\xi_1^*$ is known. Browne (1989) sets the variance of $\xi_1^*$ to one. This means that all variances and covariance must be interpreted relative to the variance of $\xi_1^*$.

## Applications of the Covariance Component Model

*To a Covariance Matrix*

For illustration we apply the CC model to three-facet data gathered by Hilton, Beaton, and Bower (1971). The data consist of $2163$ measurements of two traits (facet A), measured with two instruments (facet B) on three occasions (facet C). Details can be found in Cudeck (1988). We use the covariance matrix reported in table 4 by Cudeck (1988) to fit the CC model.

Cudeck notes that the data show large kurtosis and goodness-of-fit statistics that are based upon the assumption that the data are normally distributed may not be trusted. Following Cudeck, we provide the normal-theory generalized least-squares estimates of the parameters in $\mathbf{\Sigma}_\xi$.

$$\widehat{\mathbf{\Sigma}_\xi} = \begin{bmatrix} 92.54 & & & & \\ 6.78 & 11.94 & & & \\ 14.22 & 11.38 & 20.02 & & \\ -3.35 & -1.22 & -1.56 & 8.23 & \\ -1.44 & 0.22 & 1.93 & -5.41 & 44.04 \end{bmatrix} \quad .$$

As judged from the standardized residuals the CC model did not fit the data. The CC model has an *adjusted goodness-of-fit index (AGFI)* of $0.91$ and a *standardized root-mean-square residual (RMS)* of $0.31$. Hence, no substantive conclusions should be drawn from these parameter estimates and we merely present them to enable readers to check their results. Note that Cudeck (1988) fitted the multiplicative model that is discussed in the next section to the same covariance matrix and found that it fitted the data well.

### *To a Correlation Matrix*

Guilford's (1967) structure of intelligence model can be conceived of as a facet design for constructing intelligence tests (Fiske, 1971, p. 128). Guilford distinguished three facets. *The operation facet* refers to the subject's intellectual processing of information; *the content facet* refers to the content of the information; and *the product facet* to the form of the information. Using Guilford's facet design, Hoeks (1985) constructed eighteen tests measuring the content element semantic abilities. He selected two out of five elements of the operation facet: cognition and memory, and three out of five elements of the product facet: units, systems and transformation of information. Hence, the tests where constructed according to a $2 \times 3$ facet design. For each of the combinations, Hoeks constructed three different tests. The tests may be considered as a third facet and we treat the measures as arising from a $2 \times 3 \times 3$ facet design. The data were analyzed earlier using a standard confirmatory factor model by Hoeks, Mellenbergh and Molenaar (1989). We fit the CC model to the correlation matrix that they give in their report (see Appendix).

Following Hoeks, Mellenbergh, and Molenaar (1989) we used unweighted least-squares to fit model (10) to the correlation matrix, assuming a block-diagonal structure for $\mathbf{\Sigma}_{\xi^*}$. The CC model reproduced the observed correlation matrix well as judged from the residuals. Deviations corresponding to the third facet (different tests) showed zero variation relative to the combination of the first condition in each facet. Thus, we specified a model for two facets with three measures for each combination and found a model that fitted equally satisfactorily. This model is easily specified by deleting the columns of the third facet from the $\mathbf{A}$ matrix. Finally, we found that we could specify $\mathbf{\Sigma}_{\xi^*}$ as a diagonal matrix without visible deterioration of the fit. The final model has an AGFI of $0.99$, and a RMS of $0.051$, comparable to the values found by Hoeks, et al. (1989). Other fit indices are not reported because they require normal distributions. A LISREL script for the final model is in the Appendix. Our analysis suggests that the facet-structure suggested by Guilford does indeed hold.

## The Composite Direct Product Model

The multiplicative data model may be derived from the following multiplicative structure for the common score $\boldsymbol{\eta}$:

$$\eta_x = z(A_j)z(B_k)\cdots z(E_r) \quad . \tag{11}$$

In contrast to the CC model, it is now assumed that the true score is the product of a set of latent variables. It is clear that (11) represents a strong hypothesis; one that will not often be deemed realistic in social science applications. As illustrated in Figure 1 for two facets, different values of $z(A_i)$ give the same true score in combination with two different values of $z(B_j)$.

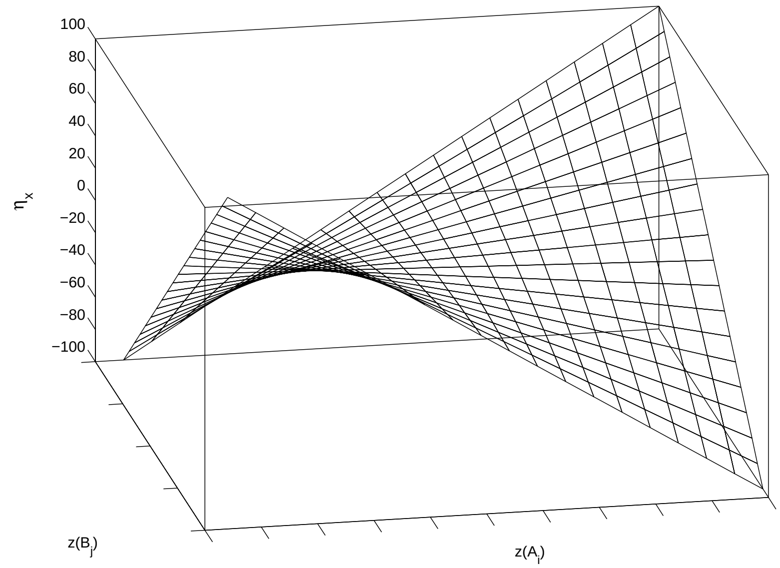


*Figure 1.* A surface plot of $\eta_x$ for different values of $z(A_i)$ and $z(B_j)$

Furthermore, it is seen that there is a unique point where the true score is zero. Needles to say that, although the latent variables in the CC and in *the composite direct product (CDP) model* are represented by the same symbol, their interpretation is quite different. In matrix notation, the data model in (11) is

$$\boldsymbol{\eta} = \mathbf{z}_A \otimes \mathbf{z}_B \otimes \mathbf{z}_C \otimes \cdots \otimes \mathbf{z}_E \quad . \tag{12}$$

As with Equation 1, Equation 12 is derived assuming that each subsequent facet is nested in the preceding facet(s).

The dispersion matrix of the common scores $\boldsymbol{\Sigma}_\eta$ is found by expanding $E\left[\boldsymbol{\eta}\boldsymbol{\eta}^T\right]$ which gives

$$\begin{aligned} \boldsymbol{\Sigma}_\eta &= E\left[\left(\mathbf{z}_A \otimes \mathbf{z}_B \otimes \cdots \otimes \mathbf{z}_E\right)\left(\mathbf{z}_A \otimes \mathbf{z}_B \otimes \cdots \otimes \mathbf{z}_E\right)^T\right] \\ &= E\left[\left(\mathbf{z}_A \otimes \mathbf{z}_B \otimes \cdots \otimes \mathbf{z}_E\right)\left(\mathbf{z}_A^T \otimes \mathbf{z}_B^T \otimes \cdots \otimes \mathbf{z}_E^T\right)\right] \quad . \end{aligned}$$

If we assume that the latent variables are normally distributed, and independent between facets, $\mathbf{\Sigma}_\eta$ has a multiplicative structure; that is,

$$\begin{aligned}\mathbf{\Sigma}_\eta &= E\left[\mathbf{z}_A\mathbf{z}_A^T \otimes \mathbf{z}_B\mathbf{z}_B^T \otimes \cdots \mathbf{z}_E\mathbf{z}_E^T\right] \\ &= \mathbf{\Sigma}_A \otimes \mathbf{\Sigma}_B \otimes \cdots \otimes \mathbf{\Sigma}_E \quad .\end{aligned}$$

It follows that

$$\mathbf{\Sigma} = \mathbf{\Sigma}_A \otimes \mathbf{\Sigma}_B \otimes \cdots \otimes \mathbf{\Sigma}_E + \mathbf{D}_u \quad . \tag{13}$$

Note that the common scores are not normally distributed under this model since the product of normally distributed variables is not normally distributed.

There is an equivalent specification of the model with standardized within-facet dispersion matrices. The true score dispersion matrix is standardized by pre- and post-multiplying the dispersion matrices by a diagonal matrix $\mathbf{D}_\eta = diag^{-\frac{1}{2}}\left[\mathbf{\Sigma}_\eta\right]$ that contains the inverse of the true score standard deviations. The multiplicative structure of $\mathbf{\Sigma}_\eta$ implies that this matrix is found to be structured as follows:

$$\mathbf{D}_\eta = \mathbf{D}_A \otimes \mathbf{D}_B \otimes \cdots \otimes \mathbf{D}_E \quad ,$$

where $\mathbf{D}_F = diag^{-\frac{1}{2}}(\mathbf{\Sigma}_F)$. Hence, the correlation matrix of the common scores, $\mathbf{P}_\eta$, is

$$\begin{aligned}\mathbf{P}_\eta &= \mathbf{D}_\eta\mathbf{\Sigma}_\eta\mathbf{D}_\eta \\ &= (\mathbf{D}_A\mathbf{\Sigma}_A\mathbf{D}_A) \otimes (\mathbf{D}_B\mathbf{\Sigma}_B\mathbf{D}_B) \otimes \cdots \otimes (\mathbf{D}_E\mathbf{\Sigma}_E\mathbf{D}_E) \\ &= \mathbf{P}_A \otimes \mathbf{P}_B \cdots \otimes \mathbf{P}_E \quad .\end{aligned}$$

The symmetric matrices $\mathbf{P}_F$ are the *within-facet correlation matrices* under the CDP model. It follows that

$$\mathbf{\Sigma} = \mathbf{D}_\eta^{-1}\left(\mathbf{P}_\eta + \mathbf{D}_\eta\mathbf{D}_u\mathbf{D}_\eta\right)\mathbf{D}_\eta^{-1} = \mathbf{D}_\eta^{-1}\left(\mathbf{P}_\eta + \mathbf{D}_U^*\right)\mathbf{D}_\eta^{-1} \quad . \tag{14}$$

The elements of the $p \times p$ diagonal matrix $\mathbf{D}_U^*$ represent ratios of unique variance to common score variance and one minus any diagonal element of $\mathbf{D}_U^*$ equals the classical test theory reliability of the corresponding measure. The matrix $\mathbf{D}_\eta^{-1}$ has a multiplicative structure since

$$\mathbf{D}_\eta^{-1} = \left(\mathbf{D}_A \otimes \mathbf{D}_B \otimes \cdots \otimes \mathbf{D}_E\right)^{-1} = \mathbf{D}_A^{-1} \otimes \mathbf{D}_B^{-1} \otimes \cdots \otimes \mathbf{D}_E^{-1} \quad .$$

The model in (14) has the same form as (10) and may be fitted using LISREL with appropriate (non-linear) constraints on $\mathbf{P}_\eta$. However, LISREL requires that each of the individual constraints be specified and this is cumbersome if there are many of them. Alternatively, the CDP model may be formulated as a CFA model (see Wothke & Browne, 1989) but this too is cumbersome when the number of facets is larger than two. Fortunately, the multiplicative model is easily fitted with the Mx program which incorporates the Kronecker product as a model operator. The following box gives a general scheme for an Mx script:

TITLE: *multiplicative model with any number of facets*
DAta NObservations=*(nr. of observations)* NInput= *(p)* NGroup=1
CM FI=*(file with sample covariance or correlation matrix)*
MATrices D DIagonal p p FREE
A STandardized a a FREE
B STandardized b b FREE
*(etc.)*
X Diagonal a a FREE
Y DIagonal b b FREE
Z DIagonal a a FREE
*(etc.)*
U DIagonal p p FREE
CO *(X@Y@· · · @Z)*(A@B@C@· · · @E + U.U)*(X@Y@· · · @Z)* /
STart *(starting values)*
Options *(here you can specify e.g., the number of iterations)*
end

When we analyze the correlation matrix we need to specify a model for the population correlation matrix. If we standardize the covariance matrix, this only affects the elements of $\mathbf{D}_F^{-1}$, which now represent the ratios of sample standard deviations to common score standard deviations.

## An Application of the Composite Direct Product Model

The Miller and Lutz (1966) data consist of the scores of 51 education students on a test designed to assess teacher's judgements about the effect of situation and instruction factors on the facilitation of student learning. The measures where constructed according to a facet design with three facets with two conditions each:

1. A: Grade level of the student. $A_1$ denotes the first grade and $A_2$ the sixth grade.

2. B: Teacher approach. $B_1$ denotes a teacher-centered approach and $B_2$ a pupil-centered approach.

3. C: Teaching method. $C_1$ denotes an approach where teaching consisted mainly of rote learning activities. $C_2$ denotes an approach in which the teacher attempts to develop pupil understanding without much emphasis on rote memorization.

The Miller-Lutz data were analyzed by Wiley, Schmidt, and Bramble (1973), and Jöreskog (1973) using the additive model, and detailed results can be found there. Note that Jöreskog (1973, p. 32) used Browne's parametrization. The CC model shows a reasonable fit to these data; $\chi^2(21) = 37.97$, $p = 0.01$, and the *Root Mean Square Error of Approximation (RMSEA)* equals $0.11$ (see Steiger & Lind, 1980 or McDonald, 1989). The results indicate that differences between the drill and the discovery methods of instruction caused most variation in the responses of the education students. Differences in teacher approach showed least variation.

We have used the covariance matrix reported by Jöreskog (1973, Table 10) to estimate the parameters of the CDP model. The model did not fit the data as judged from the chi-square statistic ($\chi^2(19) = 49.4$, $p < 0.001$, $RMSEA = 0.18$) as well as the residuals. For illustration we give some of the results here:

$$\mathbf{P}_A = \begin{bmatrix} 1 & \\ 0.89\,(0.81 - 0.95) & 1 \end{bmatrix}, \mathbf{P}_B = \begin{bmatrix} 1 & \\ 0.92\,(0.84 - 0.98) & 1 \end{bmatrix}$$

and

$$\mathbf{P}_C = \begin{bmatrix} 1 & \\ 0.42\,(0.19 - 0.61) & 1 \end{bmatrix}$$

Within brackets are $95$-percent confidence intervals (Neale & Miller, 1997). The results confirm the conclusion that, in the view of the coming teachers, differences in teaching method are more important than differences on any other facet. On the other hand, there is no substantive reason to support the CDP model for these data.

## Concluding Remarks

In this paper we have extended models that were conceived for the analysis of MTMM data. We have demonstrated how the models are derived from the model for the observations and how they can be fitted using the LISREL or the Mx program. Note that Mx can handle all models that have been discussed. A minor disadvantage of Mx program is that it uses numerical derivatives which may make the optimization algorithm less stable, sometimes.

Models that are similar to the CDP model are described by Swain (1975) and Verhees and Wansbeek (1990) and the Mx script described above is easily adapted to fit these models. It is possible and indeed not difficult to formulate *hybrid* models combining an additive specification of some facets and a multiplicative specification for others. Such models are not difficult to fit using Mx. However, unless there is a strong theoretical interest in such models, fitting them would merely be an exercise in SEM. This brings us to an important point. To wit, although the CDP model has been found useful to describe MTMM correlation matrices, it represents a strong hypothesis on the data. We find it somewhat disturbing that the vast majority of the applications of the CDP model to MTMM matrices *that we know* provide no substantive arguments for use of the model. An exception being, for instance, Bagozzi, Yi and Phillips (1991). Even studies where multiplicative and additive models are compared (e.g., Hernández Baeza & González Romá, 2002) focus almost exclusively on the relative fit of the models. At most, authors (e.g., Cudeck, 1988, p. 141) refer to the work by Campbell and O'Connell (1967; 1982) who observed that for some MTMM correlation matrices, inter-trait correlations are attenuated by a multiplicative constant (smaller in magnitude than unity), when different methods are used.

In closing, we mention two topics for future research. First, it is necessary to establish the identifiability of the (reparameterized) CC and CDP model. Although we believe these models to be identifiable there is no general proof available that they are identifiable for any number of facets. Second, we would like to have ways to perform exploratory analysis on the within-facet covariance (or correlation) matrices. A suggestion is to use a model incorporating principal components. Such model have been considered (for two-facets) by Flury and Neuenschwander (1995). Dolan, Bechger, and Molenaar (1999) suggest how these model can be fitted in a SEM framework.

# REFERENCES

Bagozzi, R. P., Yi, Y. & Phillips, L. W. (1991). Assessing construct validity in organizational research. *Administrative Science Quarterly*, **36**, 421-458.

Bagozzi, R. P., Yi, Y., & Nassen, K. D. (1999). Representation of measurement error in marketing variables: Review of approaches and extension to three-facet designs. *Journal of Econometrics*, **89**, 393-421.

Bechger, T. M., Verhelst, N. D., & Verstralen, H. H. F. M. (2001). Identifiability of nonlinear logistic test models. *Psychometrika*, *66*, 357-372

Bock, R. D., & Bargmann, R. E. (1966). Analysis of covariance structures. *Psychometrika*, **31**(4), 507-534.

Bollen, K. A. (1989). *Structural Equations with Latent Variables.* New York: Wiley.

Browne, M. W. (1970). *Analysis of covariance structures.* Paper presented at the annual conference of the South African Statistical Association.

Browne, M. W. (1984). The decomposition of multitrait-multimethod matrices. *British Journal of Mathematical and Statistical Psychology*, **37**, 1-21.

Browne, M. W. (1989). Relationships between an additive model and a multiplicative model for multitrait-multimethod matrices. In R. Coppi & S. Bolasco. (Eds.).*Multiway Data Analysis*. New-York: Elsevier.

Browne, M. W. (1993). Models for Multitrait Multimethod Matrices. in R Steyer, K. F. Wenders & Widaman, K.F. (Eds.) *Psychometric Methodology. Proceedings of the 7th European Meeting of the Psychometric Society in Trier*. (pp. 61-73) New York : Fisher.

Browne, M. W., & Strydom, H. F. (1997). Non-iterative fitting of the direct product model for multitrait-multimethod matrices. In M. Berkane (Editor). *Latent variable modeling and applications to causality*. New-York: Springer.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin*, **56**, 81-105.

Campbell, D. T., & O'Connell, E. J. (1967). Method factors in multitrait-multimethod matrices: Multiplicative rather than additive? *Multivariate Behavioral Research*, **2**, 409-426.

Campbell, D. T., & O'Connell, E. J. (1982). Methods as diluting trait relationships rather than adding irrelevant systematic variance. In D. Brinberg and L.H. Kidder (Eds.) *Forms of validity in research. New directions for methodology of social and behavioral science*. Vol. 12.

Cudeck, R. (1988). Multiplicative Models and MTMM matrices. *Journal of Educational Statistics*, **13**(2), 131-147.

Cudeck, R. (1989). Analysis of correlation matrices using covariance structure models. *Psychological Bulletin*, **105**(2), 317-327.

Dolan, C. V. , Bechger, T. M., & Molenaar, P. C. M. (1999). Fitting models incorporating principal components using LISREL 8. *Structural Equation Modeling*, **6**(3), 233-261.

Fiske, D. W. (1971). *Measuring the concepts of intelligence*. Chicago: Aldine.

Flury, B. D., & Neuenschwander, B. E. (1995). Principal component models for patterned covariance matrices with applications to canonical correlation analysis of several sets of variables. Chapter 5 In W.J. Krzanowski (Ed.) *Recent Advances in Descriptive Multivariate Analysis*, Oxford: Oxford University Press.

Guilford, J. P. (1967). *The nature of human intelligence*. New-York: McGraw-Hill.

Hernández Baeza, A., & González Romá, V. (2002). Analysis of Multitrait-Multioccasion Data: Additive versus Multiplicative Models. *Multivariate Behavioral Research*, **37**(1), 59-87.

Hilton, T. L., Beaton, A. E., & Bower, C. P. (1971). *Stability and instability in academic growth: A compilation of longitudinal data*. Final report, USOE, Research No 0-0140. Princeton, NJ: ETS.

Hoeks, J. (1985). *Vaardigheden in begrijpend lezen*. [Abilities in reading comprehension]. Unpublished dissertation, University of Amsterdam.

Hoeks, J., Mellenbergh, G. J., & Molenaar, P. C. M. (1989) *Fitting linear models to semantic tests constructed according to Guilford's facet design.* University of Amsterdam.

Jöreskog, K. G. (1973). Analyzing psychological data by structural analysis of covariance matrices. In D.H. Krantz, R.D. Luce., R.C. Atkinson & P. Suppes (Eds.) *Measurement, psychophysics and neural information processing.* Volume II of Contemporary developments in mathematical Psychology. San Fransisco: W.H. Freeman and Company.

Jöreskog, K. G. , & Sörbom, D. (1996). *LISREL8 User's reference Guide*. SSI.

McDonald, R. P. (1989). An index of goodness-of-fit based on noncentrality. *Journal of Classification*, *6*, 97-103.

Neale, M. C. , Boker, S. M. , Xie, G., & Maes, H. (2002). *Mx: Statistical modelling*. Virginia Institute for Psychiatric and Behavioral Genetics, Virgina Commonwealth University, Department of Psychiatry.

Neale, M. C. , & Miller, M. B. (1997). The use of likelihood-based confidence intervals in genetic models. *Behaviour Genetics*, **20**, 287-298.

Pringle, R. M., & Rayner, A. A. (1971). *Generalized inverse matrices with applications to statistics.* London: Griffin.

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

Swain, A. J. (1975). *Analysis of Parametric Structures for Covariance Matrices*. Unpublished PhD thesis. University of Adelaide.

Verhees, J. , & Wansbeek, T. J. (1990). A multimode direct product model for covariance structure analysis. *British Journal of Mathematical and Statistical Psychology*, **43**, 231-240.

Wiley, D., Schmidt, W. H., & Bramble, W. J. (1973). Studies of a class of covariance structure models. *Journal of the American statistical association*, **68**(342), 317-323.

Wothke, W. (1984). *The estimation of trait and method components in multitrait multimethod measurement*. University of Chigago, Department of Behavioral Science: Unpublished dissertation.

Wothke, W. (1988). *Identification conditions for scale-free covariance component models.* Paper Presented at the Annual Meeting of the Psychometric Society, Los Angeles, June 26-29.

Wothke, W. (1996). Models for multitrait-multimethod matrix analysis. Chapter 2 in *Advanced techniques for structural equation modeling*, Edited by R.E. Schumacker and G.A. Marcoulides. New-York: Lawrence Erlbaum Associates.

Wothke, W., & Browne, M. W. (1990). The direct product model for the MTMM matrix parameterized as a second order factor analysis model. *Psychometrika*, **55**, 255-262.

# Appendix

```
TItle Hoeks data CC model on correlations
Da No=620 Ni=18
KM SY
1.000
0.502 1.0000
0.512 0.475 1.0000
0.419 0.313 0.419 1.000
0.457 0.457 0.430 0.385 1.000
0.481 0.401 0.479 0.431 0.405 1.000
0.568 0.519 0.530 0.501 0.553 0.497 1.000
0.456 0.396 0.494 0.512 0.491 0.449 0.637 1.000
0.527 0.470 0.444 0.453 0.490 0.491 0.680 0.675 1.000
0.485 0.433 0.418 0.373 0.415 0.361 0.544 0.425 0.533 1.000
0.426 0.388 0.392 0.384 0.407 0.339 0.445 0.413 0.436 0.441
1.000
0.293 0.282 0.306 0.259 0.262 0.253 0.290 0.222 0.271 0.273
0.243 1.000
0.573 0.508 0.577 0.487 0.482 0.516 0.642 0.555 0.558 0.516
0.404 0.303 1.000
0.516 0.389 0.465 0.390 0.420 0.436 0.517 0.456 0.461 0.434
0.348 0.286 0.582 1.000
0.497 0.398 0.410 0.445 0.452 0.483 0.616 0.572 0.598 0.552
0.401 0.247 0.601 0.494 1.000
0.160 0.101 0.212 0.151 0.093 0.256 0.152 0.196 0.156 0.186
0.117 0.137 0.166 0.119 0.159 1.000
0.206 0.106 0.140 0.162 0.204 0.239 0.131 0.103 0.073 0.170
0.178 0.333 0.215 0.288 0.207 0.106 1.000
0.251 0.168 0.226 0.233 0.250 0.378 0.248 0.229 0.221 0.250
0.169 0.272 0.293 0.267 0.276 0.440 0.290 1.000
mo ny=18 ne=18 nk=4 ps=ze,fi ly=di,fr ga=fu,fi ph=sy,fr
MA gamma
1 0 0 0 !0 0
1 0 0 0 !1 0
1 0 0 0 !0 1
1 0 0 0 !0 0
1 0 0 0 !1 0
1 0 0 0 !0 1
1 0 1 0 !0 0
```

```
1 0 1 0 !1 0
1 0 1 0 !0 1
1 1 1 0 !0 0
1 1 1 0 !1 0
1 1 1 0 !0 1
1 1 0 1 !0 0
1 1 0 1 !1 0
1 1 0 1 !0 1
1 1 0 1 !0 0
1 1 0 1 !1 0
1 1 0 1 !0 1
pa ph
0
0 1
0 0 1
0 0 0 1
!0 0 0 0 1
!0 0 0 0 1 1
va 1 ph(1,1)
st 0.3 all
ou se rs ULS it=40000
```

Note that anything after a ! is ignored by LISREL. We have kept it here to illustrate how the script was changed from the first to the final analysis.