ON THE VALIDITY OF
THE INDETERMINATE LATENT VARIABLES IN THE LISREL MODEL

Giorgio Vittadini

Istituto di Statistica Marcello Boldrini
Facoltà di Scienze Politiche
Università degli Studi di Milano
Via Conservatorio 7, Milano, Italia, 20159

Key Words and Phrases: Latent variables; Identification; Vectorial spaces; Isomorphism; Adjunctive variables; Criteria uncorrelated with latent variables and perfectly correlated with observed variables.

ABSTRACT

The language of abstract vector spaces is used to show that the latent variables and errors of the Lisrel model can always be constructed so as to predict any criterion perfectly, including all those that are entirely uncorrelated with the observed variables.

1. INTRODUCTION

In factor analysis two types of indeterminacy problems are considered. The first problem is known as the problem of

2. THE LISREL MODEL

The Lisrel model considers vectors of random variables:

$$\mathbf{z} = (z_1, z_2, \ldots, z_p)'$$

$$\mathbf{t} = (t_1, t_2, \ldots, t_q)'$$

of latent dependent and independent variables respectively in the following system of linear structural relations:

$$\mathbf{z} = \mathbf{Az} + \mathbf{Pt} + \mathbf{g}$$

where \( \mathbf{g} \) (fx1) is a vector of residuals (errors in equations) and \( \mathbf{P} \) (fx1) and \( \mathbf{A} \) (gxg) are coefficient matrices (\( \mathbf{A} \) has zeroes in the main diagonal and \( \mathbf{1} - \mathbf{A} \) is non singular).

There are assumed linear relations between the vectors \( \mathbf{z} \) and \( \mathbf{t} \) and the observed vectors \( \mathbf{x} = (x_1, x_2, \ldots, x_p)' \) and \( \mathbf{y} = (y_1, y_2, \ldots, y_q)' \):

$$\mathbf{x} = \mathbf{Bz} + \mathbf{d}$$

$$\mathbf{y} = \mathbf{Ct} + \mathbf{e}$$

where \( \mathbf{B} \) (qxf) and \( \mathbf{C} \) (rxg) are regression matrices of \( \mathbf{x} \) on \( \mathbf{z} \) and of \( \mathbf{y} \) on \( \mathbf{t} \); \( \mathbf{d} \) (qx1) and \( \mathbf{e} \) (rx1) are two sets of error of measurement (error in variables) in \( \mathbf{z} \) and \( \mathbf{y} \) respectively.

INDETERMINATE LATENT VARIABLES

We assume that:

$$f \leq q ; r \leq g$$

$$E(\mathbf{1}) = 0; E(\mathbf{g}) = 0; E(\mathbf{d}) = 0; E(\mathbf{e}) = 0;$$

$$E(\mathbf{t}'\mathbf{g}) = 0; E(\mathbf{t}'\mathbf{d}) = 0; E(\mathbf{t}'\mathbf{e}) = 0; E(\mathbf{g}'\mathbf{d}) = 0; E(\mathbf{g}'\mathbf{e}) = 0; E(\mathbf{d}'\mathbf{e}) = 0.$$

Because we are dealing here with the indeterminacy problem of latent variables and errors, we assume that the parameter matrices \( \mathbf{A, P, B, C} \) are known.

Our problem is now how we can determine the latent variables having the observations of \( n \) independent drawings of the random vectors \( \mathbf{x} \) and \( \mathbf{y} \).

Let us indicate by \( \mathbf{X} \) and \( \mathbf{Y} \) the (qxn) and (rxn) matrices of observed scores of \( \mathbf{x} \) and \( \mathbf{y} \); by \( \mathbf{Z} \) and \( \mathbf{T} \) the (fxn) and (gxn) scores matrices of \( \mathbf{z} \) and \( \mathbf{t} \); by \( \mathbf{G} \) the (pxn) score matrix of \( \mathbf{g} \); by \( \mathbf{D} \) and \( \mathbf{E} \) the (qxn) and (rxn) score matrices of \( \mathbf{d} \) and \( \mathbf{e} \), assuming the corresponding properties in (4) and (5) for the scores.

The Lisrel model can be written:

$$\mathbf{Z} = \mathbf{AZ} + \mathbf{PT} + \mathbf{G}$$

$$\mathbf{X} = \mathbf{BZ} + \mathbf{D}$$

$$\mathbf{Y} = \mathbf{CG} + \mathbf{E}$$

Let \( \mathbf{S}_x (\mathbf{q} \times \mathbf{q}) \), \( \mathbf{S}_y (\mathbf{p} \times \mathbf{p}) \), \( \mathbf{S}_d (\mathbf{q} \times \mathbf{r}) \), \( \mathbf{S}_e (\mathbf{r} \times \mathbf{r}) \) be the covariance matrices of \( \mathbf{Z, G, D, E} \) respectively, and:

$$\mathbf{S} = \begin{bmatrix} \mathbf{S}_x & 0 \\ 0 & \mathbf{S}_g \\ 0 & \mathbf{S}_d \\ \mathbf{S}_e \end{bmatrix}$$

Then it follows from the above assumptions that the covariance matrix \( \mathbf{S}_{j} (\mathbf{q} + \mathbf{r} + \mathbf{r}) \) of \( J = (\mathbf{X}, \mathbf{Y}) \) is:

$$\mathbf{S}_{j} = \mathbf{F} \mathbf{S}_{j} \mathbf{F}^{+}.$$
3. BASIC NOTATION

Let $\mathbb{R}^n$ be the vector space of the n-dimensional vectors with real valued components and $V^j_n$ be the vector space of all mean centered row vectors of n components. We write $V^j_n \subset \mathbb{R}^n$ to indicate that $V^j_n$ is a proper subspace of $\mathbb{R}^n$.

If $l_1, l_2, \ldots, l_g$ are the g rows of the matrix $T$ $(g \times n)$ we denote its row space (the set of all linear combinations of its rows) by $V^g_t$.

For $V^j_t \subset V^j_t$ and $V^k_t \subset V^k_t$ we call $V^j_t \oplus V^k_t$ the direct sum of $V^j_t$ and $V^k_t$ and we write:

$$V^j_t = V^j_t \oplus V^k_t$$

if any vector $x \in V^j_t$ can be uniquely decomposed as the sum $x = x^j_t + x^k_t$ with $x^j_t \in V^j_t$ and $x^k_t \in V^k_t$ vectors uniquely defined.

For:

$$V^j_t = V^j_{t_1} \oplus V^j_{t_2}$$

we call $V^j_{t_2}$ an orthogonal complement of $V^j_t$ relative to $V^j_{t_1}$ and we write:

$$V^j_t = V^j_{t_1} \oplus V^j_{t_2}$$

if any $x \in V^j_t$ can be written as $x = x^j_{t_1} + x^j_{t_2}$ with $x^j_{t_1} \in V^j_{t_1}$ and $x^j_{t_2} \in V^j_{t_2}$.

The idempotent and symmetrical matrix $P^j_t$:

$$P^j_t = T^j_t (T^j_t)^{-1}$$

is the orthogonal projector onto $V^j_t$; the vector:

$$x = x^j_t + x$$

with $x^j_t \in V^j_t$, $x \in V^j_t$ in $V^j_t$ is the orthogonal projection on $V^j_t$ along $V^j_t$.

4. THE INDETERMINACY OF THE LISREL MODEL

The Lisrel model (5)-(10) can be written in terms of language of abstract vector spaces.

The rows of the matrix $J$ are contained in $V^j_n$ and span the subspace $V^j_n$ of dimension $q+r$ in $V^j_n$.

The rows of the matrix of latent variables $T$ span a subspace $V^j_n$ of dimension $g$ in $V^j_n$; the rows of the matrix of errors in equations $G$ span a subspace $V^j_n$ of dimension $l + (q+r+f+g)$.

Due to assumption (4) $V^j_n \oplus V^j_n \oplus V^j_n \oplus V^j_n$ are mutually orthogonal:

$$V^j_n = V^j_t \oplus V^j_g \oplus V^j_e \oplus V^j_e$$

(12) implies that:

$$P^j_1 = P^j_t + P^j_g + P^j_e$$

with $P^j_1, P^j_t, P^j_g, P^j_e$ orthogonal projectors onto $V^j_n, V^j_n, V^j_n, V^j_n$ respectively. Since:

$$V^j_n \subset V^j_n$$

there exists the complementary subspace of $V^j_n$ in $V^j_n$ which we denote by $V^j_n$. Hence:

$$V^j_n = V^j_1 \oplus V^j_n$$

(15) and,

$$P^j_n = P^j_t + P^j_g + P^j_e$$

(16) with $P^j_n$ orthogonal projector onto $V^j_n$.

Since the rows of the matrix $J$ are linear combinations of the rows of the matrices $T, G, F, D$, we have:

$$V^j_n \subset V^j_n$$

and:

$$J P^j_n = 0$$

(18) and:

$$J P^j_n = J$$

(19)

Hence, it results:

$$J = J P^j_n + J P^j_n + J P^j_n + J P^j_n$$

(20)

To determine Lisrel solutions we have to decompose the space $V^j_n$ in the direct sum of the spaces $V^j_t \oplus V^j_g \oplus V^j_e$; then decompose the row vectors of $J$ in the sum of their projections on the spaces $V^j_t, V^j_t, V^j_e$. As long as $n-1$ exceeds 1 the decomposition of the space $V^j_n$ and of row-vectors $J$ are not univocal. Thus, there are infinite sets of matrices $(T, G, D, E)$ which satisfy Lisrel model (4).
5. ADDITIONAL VARIABLES AND INDETERMINACY OF THE LISREL MODEL

The indeterminacy of the Lisrel model has relevant consequences. On partialling out the observed scores $J$ from the latent variables $T$ and the errors in equations $G$ the partial covariance matrices cannot vanish. By (8)-(11) we have:

$$
S(t/j) S_t = S(t) C'S_S^{-1} C S - S P'(I-A')^{-1} B'S_S^{-1} B(I-A)^{-1} P S_t
$$

with (21) and (22) positive definite matrices.

We introduce a set of new $f \times g$ additional deviation scores $M$ as:

$$
K = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix} \quad KJ = 0 \quad KK' = I
$$

We can prove that the latent variables $T$, the errors in equations $G$ and the errors in variables $D$ and $E$ can be expressed as function of observed and additional variables:

$$
\begin{bmatrix} T \\ G \end{bmatrix} = \begin{bmatrix} S_t F_t' S_S^{-1} : Q_t \\ S G F' G^{-1} : Q \end{bmatrix} \begin{bmatrix} J \\ K \end{bmatrix}
$$

with:

$$
F_t' = \begin{bmatrix} P'(I-A')^{-1} B' : C' \end{bmatrix}
$$

$$
G' = \begin{bmatrix} (I-A')^{-1} B : 0 \end{bmatrix}
$$

and:

$$
L = \Gamma W
$$

with:

$$
\begin{bmatrix} T \\ G \\ D \\ E \end{bmatrix} = \begin{bmatrix} X \\ Y \\ K \end{bmatrix}
$$

where $C$ is the inverse of the correlation matrix of observed variables.

We can also prove, that, after the introduction of additional variables $K$ the partial correlation matrices (21) and (22) vanish.

By the incorrelation of additional variables $K$ with the variables $J$ we have:

$$
V_j = V_j^k
$$

and if

$$
\text{dim } V_j^k = n - 1 - q - r = (\text{dim } V = f + g)
$$

then the complement space to $V$ as regard $V_j^k$ exists:

$$
V_j^k = V_k \oplus V_j^k
$$

As long as $n - 1$ exceeds $q + r + f + g$ there is leeway for the definition of $K$ and therefore $T$ and $G$.

6. FIRST THEOREM

Schönemann and Steiger (1978) have demonstrated that the indeterminate orthogonal factors in factor analysis can always be constructed so as to predict any criterion perfectly, including all those that are entirely uncorrelated with the observed variables. We prove that similar implications are valid also for the non-orthogonal latent variables and errors of the Lisrel model. The first theorem can be written in the following form:
correlated with the observed variables \( J \) but completely unpredictable (in a multiple-regression sense) from the latent variables \( T \) under all the choices of \( K \).

b) If \( n > q + r + f + g \) and \( m + r > f + g \), then a criterion \( \lambda_1 \) exists, perfectly correlated with the observed variables but completely unpredictable from the \( f + g \) latent variables and errors in equations \( T \) and \( G \).

**PROOF**

Let \( V_{k,t} \), \( V_{t,j} \), \( V_{t} \) be the spaces generated by the rows of \( T_{k}^{-1} F_{t}^T S_{t,j} \), \( Q_{t} \), \( K \), \( K \) respectively. From (24) we find that:

\[
V_{t} = V_{t,j} \oplus V_{k,l} C (V_{t,j} \ominus V_{k,l})
\]

Let \( \lambda E V_{t,j} \) be a normalized criterion.

Being \( V_{t,j} \subset V_{t} \)

from (26) and (29) we have:

\[
V_{t,j} \perp (V_{t,j} \ominus V_{t}) = V_{t}
\]

and, therefore:

\[
\lambda \perp V_{t}
\]

The squared multiple correlation of \( \lambda \) regressed on any set of linear independent predictors \( \frac{1}{n} \sum_{j=1}^{n} t_{j} \) in \( V_{t} \) is (Searle 1966):

\[
R_{\lambda t}^2 = \frac{\lambda^T (\lambda \ominus V_{k})}{\lambda^T \lambda}
\]

where \( \lambda_{t} \) is orthogonal projector in \( V_{t} \).

From the projector properties (Takeuchi 1982p.28) through (31) we have the result of:

\[
\frac{1}{n} \sum_{j=1}^{n} t_{j} = 0
\]

and, therefore:

\[
R_{\lambda t}^2 = 0
\]

Again for the orthogonal projectors properties:

\[
\lambda \perp V_{t,j} \iff \perp V_{t,j} = \lambda_{t,j}
\]

Therefore the squared multiple correlation of \( \lambda \) regressed on any set of variables of the space \( V_{t} \) is:

\[
R_{\lambda t,j}^2 = \frac{1}{n} \sum_{j=1}^{n} t_{j} / t_{j} = 1
\]

Remembering that \( V_{t,j} \subset V_{t} \)

\[
R_{\lambda t,j}^2 = 1 - R_{\lambda t,j}^2
\]

**7. SECOND THEOREM**

If \( n > q + r + f + g \) and the vectors of latent variables \( T \) are linearly independent, then a criterion \( \lambda \) is predictable (in a multiple regression sense) from the latent variables \( T \) and completely uncorrelated with the observed scores \( J \) independently of the choice of \( K \).

**PROOF**

Let \( V_{k} \), \( V_{1} \), \( V_{m} \) be the spaces generated by the rows of the matrices \( K \), \( L \), \( M \) defined in (23). Let \( u \) be a criterion such that:

\[
u \in V_{1} \quad \text{or} \quad u \in V_{m} \quad \text{or} \quad u \in V_{k} = (V_{1} \oplus V_{m})
\]

and:

\[
R_{\lambda k}^2 = 0
\]

We can use the vector \( u \) and \( f + g \) nonnull vectors such that:

\[
K = \begin{bmatrix}
u \\
b_{2} \\
b_{3} \\
\vdots \\
b_{f+g-1}
\end{bmatrix}
\]

\[
K = (f + g) - I
\]

\[
\frac{1}{n} u K' = 1 = (1, 0, 0, \ldots, 0)
\]

to construct a basis for \( V_{k} \).

The square multiple correlation of \( u \) with \( T \) is:
INDETERMINATE LATENT VARIABLES

\[ h \in V^1_n = V^1_j + V^1_j \]

h can only be written as:

\[ h = g + k \]

If \( k = 0 \) we have for (23), (24), (25):

\[ R^2_{X} = 1 = R^2_{Y} \]

For the isomorphism:

\[ R^2_{X} = R^2_{Y} = 1 \]

On the contrary, if \( k \neq 0 \), we construct the vector:

\[ \mathbf{k}^* = \mathbf{k} (n / k \mathbf{k}^*) \]

thus:

\[ (k^*k^*)/n = 1 \]

Like \( k \) also \( k^* \in V^1_j \) : this one can be used to construct a basis \( K \) for \( V^1_j \) with \( f^*g \) other vectors \( b \) such as:

\[ k^* = 0 \]

As \( V^1_j \) contains \( g \) and \( V^1_j \) contains \( k \), we have:

\[ h \in (V^1_j) \cap V^1_j \]

and:

\[ R^2_{X} = 1 \]

To conclude for the isomorphism we have:

\[ R^2_{X} = R^2_{Y} = 1 \]

CONCLUSION

In the LISREL model on partialling out the observed scores from the latent variables and errors in equations the partial covariance matrices cannot vanish. We can get round the problem by introducing a set of dependent adjunctive variables into the model. But we can demonstrate that, also in this case, the LISREL model solutions remain indeterminate. This 'indeterminacy' of the solutions of LISREL yields some relevant consequences.

Criteria completely uncorrelated with observed scores, are predictable (in a multiple regression sense) from the latent variables and errors in equation. On the other hand, criteria perfectly correlated with observed scores, are completely uncorrelated with the latent variables and errors in equation.
Moreover a criterion exists, perfectly correlated with latent variables and errors, and completely uncorrelated with the observed scores.

**BIBLIOGRAPHY**


Received by Editorial Board member September, 1987; Revised November, 1987.

Recommended by A. M. Kshirsagar.