

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

identification of parameters of the model (Reiersol 1950, Howe 1955, Anderson Rubin 1956, Lawley Maxwell 1963, Jöreskog 1967, Haagen 1983); the second problem, the indeterminacy of factor scores, concerns the determination of factor scores (Piaggio 1931, Ledermann 1938, Kestelman 1952, Guttman 1955, Heermann 1964, Heermann 1966, Schönemann 1971, Schönemann Wang 1972, Schönemann Steiger 1976, Schönemann Steiger 1978, Schönemann Haagen 1986, Haagen 1986). 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. In the following paper we shall show that similar implications are valid also for the non orthogonal latent variables and errors of the Lisrel model (Jöreskog 1973, 1977, 1981, 1982, 1982a, Jöreskog Sörbom 1978).

2. THE LISREL MODEL

The Lisrel model considers vectors of random variables:

$$\underline{z} = (z_1, z_2, \dots, z_r)'; \quad \underline{t} = (t_1, t_2, \dots, t_g)'$$

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

$$\underline{z} = \underline{A}\underline{z} + \underline{P}\underline{t} + \underline{g} \quad (1)$$

where $\underline{g}$ (fxl) is a vector of residuals (errors in equations) and $P(fxg)$ and $A(gxg)$ are coefficient matrices (A has zeroes in the main diagonal and $I-A$ is non singular).

There are assumed linear relations between the vectors $\underline{z}$ and $\underline{t}$ and the observed vectors $\underline{x} = (x_1, x_2, \dots, x_q)'$ and $\underline{y} = (y_1, y_2, \dots, y_r)'$

$$\underline{x} = \underline{B}\underline{z} + \underline{d} \quad (2)$$

$$\underline{y} = \underline{C}\underline{t} + \underline{e} \quad (3)$$

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

We assume that:

$$f \leq q; \quad r \leq g \quad (4)$$

$$E(\underline{t})=0; \quad E(\underline{g})=0; \quad E(\underline{d})=0; \quad E(\underline{e})=0; \quad (5)$$

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

Because we are dealing here with the indeterminacy problem of latent variables and errors, we assume that the parameters matrices 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 $\underline{x}$ and $\underline{y}$.

Let us indicate by X and Y the ($q \times n$) and ($r \times n$) matrices of observed scores of $\underline{x}$ and $\underline{y}$; by Z and T the ($f \times n$) and ($g \times n$) scores matrices of $\underline{z}$ and $\underline{t}$; by G the ($p \times n$) score matrix of $\underline{g}$; by D and E the ($q \times n$) and ($r \times n$) score matrices of $\underline{d}$ and $\underline{e}$, assuming the corresponding properties in (4) and (5) for the scores.

The Lisrel model can be written:

$$Z = AZ + PT + G \quad (6)$$

$$X = BZ + D \quad (7)$$

$$Y = CG + E \quad (8)$$

Let $S_t(qxq)$, $S_g(pxp)$, $S_d(qxq)$, $S_e(rxr)$ be the covariance matrices of T, G, D, E respectively, and:

$$F = \begin{bmatrix} B(I-A)^{-1}P & B(I-A)^{-1} & I_q & 0 \\ C & 0 & 0 & I_r \end{bmatrix} \quad (9)$$

$$S = \begin{bmatrix} S_t & 0 & & \\ & S_g & & \\ & & S_d & \\ & & 0 & S_e \end{bmatrix} \quad (10)$$

Then it follows from the above assumptions that the covariance matrix $S_j(q+rx+q+r)$ of $J' = (X, Y)$ is:

$$S_j = F' S_j F \quad (11)$$

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:

$$S_{(t/j)} S_t - S_t C S_j^{-1} C S_t - S P' (I-A)^{-1} B' S_j^{-1} B (I-A)^{-1} P S_t \quad (21)$$

$$S_{(g/j)} S_g - S_g B' S_j^{-1} B S_g \quad (22)$$

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

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

$$K = \begin{bmatrix} K_1 \\ K_2 \end{bmatrix}; \quad KJ' = 0; \quad KK' = I \quad (23)$$

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:

$$\begin{bmatrix} T \\ G \end{bmatrix} = \begin{bmatrix} S_t F_t' & S_j^{-1}; & Q_t & x \\ S_g F_g' & S_j^{-1}; & Q_g & k \end{bmatrix} \begin{bmatrix} J \\ x \\ K \end{bmatrix} \quad (24)$$

with:

$$F_t' = [P' (I-A')^{-1} B'; C'];$$

$$F_g' = [(I-A')^{-1} B'; 0]$$

and:

$$L = \Gamma W \quad (25)$$

with:

$$L = \begin{bmatrix} T \\ G \\ D \\ E \end{bmatrix}; \quad W = \begin{bmatrix} X \\ Y \\ K \\ 1 \\ K_2 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} S_t P' (I-A') & B' S_{ja} & S_t C' S_{jb} & Q_1 & Q_2 \\ S_g (I-A') & B' S_{ja} & 0 & \Phi_1 & \Phi_2 \\ S_d S_{ja} & 0 & B (I-A)^{-1} (PQ_1 + \Phi_1) & B (I-A) & (PQ_2 + \Phi_2) \\ 0 & S S_{eja} & C_{11} & C_{12} & C_{22} \end{bmatrix}$$

with:

$$S_j^{-1} = \begin{bmatrix} S_{ja} & S_{jb} \\ S_{ja} & S_{jb} \end{bmatrix} \text{ inverse of the correlation matrix of observed variables}$$

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:

$$V_j^{-1} V_k \quad (26)$$

and if

$$(\dim V_j^{-1} = n-1-q-r) > (\dim V_k = f+g) \quad (27)$$

then the complement space to V_k as regard V_j^{-1} exists:

$$V_j^{-1} = V_k \oplus V_k^\perp \quad (28)$$

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 $\bar{I}_1$ 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 $V_{t1}, V_{t2}, \dots, V_{tk}$ be the spaces generated by the rows of $S_t P_t S_t^{-1} J$, $Q_t K, K$ respectively. From (24) we find that:

$$V_t = V_{tj} \oplus V_{kl} C(V_{tj} \oplus V_k) \quad (29)$$

Let $\bar{I} \in V_{tj}^\perp$ be a normalized criterion.

Being $V_{tj} \subset V_j$

from (26) and (29) we have :

$$V_{tj}^\perp \perp (V_{tj} \oplus V_k) = V_t \quad (30)$$

and , therefore:

$$\bar{I} \perp V_t \quad (31)$$

The squared multiple correlation of $\bar{I}$ regressed on any set of linear independent predictors $t_1, t_2, \dots, t_g$ in V_k is (Searle 1966):

$$R_{\bar{I}t}^2 = \bar{I} P_t \bar{I} \quad (32)$$

where P_t is orthogonal projector in V_t .

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

$$\bar{I} P_t = 0 \quad (33)$$

and ,therefore:

$$R_{\bar{I}t}^2 = 0 \quad (34)$$

Again for the orthogonal projectors properties:

$$\bar{I} \in V_{tj}^\perp \text{ iff } \bar{I} P_{tj}^\perp = P_{tj}^\perp \bar{I} \quad (35)$$

Therefore the squared multiple correlation of $\bar{I}$ regressed on any set of variables of the space V_k is:

$$R_{\bar{I}tj}^2 = \frac{1}{n} \bar{I} P_{tj}^\perp P_{tj}^\perp \bar{I} = 1 \quad (36)$$

Remembering that $V_{tj} \subset V_j$

$$R_{\bar{I}tj}^2 = 1 = R_{\bar{I}j}^2 \quad (37)$$

For the errors in equations G the demonstration is similar. For the latent variables and errors in equation $f+g$, the demonstration is identical.

On the contrary, if the number of the observed variables J is identical to the number of the latent variables and the errors in equation , by (24) we deduce that the space generated by the observed variables is isomorphic to the space generated by the latent variables and the errors in equation. Therefore it is impossible to define a criterion $\bar{I} \perp V_t$ such that $R_{\bar{I}j}^2 = 1$.

7. SECOND THEOREM

If $n > q+r+f+g$ and the vectors of latent variables T are linearly independent, then a criterion $\bar{u}$ 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 $V_k, V_1, \dots, V_m$ be the spaces generated by the rows of the matrices K, L, M defined in (23). Let $\bar{u}$ be a criterion such that:

$$\bar{u} \in V_1 \text{ or } \bar{u} \in V_m \text{ or } \bar{u} \in V_k = (V_1 \oplus V_m) \quad (38)$$

and:

$$R_{\bar{u}j}^2 = 0 \quad (39)$$

We can use the vector $\bar{u}$ and $f+g-1$ nonnull vectors such that:

$$K = \begin{bmatrix} \bar{u} \\ b_2 \\ \vdots \\ b_j \\ \vdots \\ b_{f+g-1} \end{bmatrix} \quad KK' = I_{f+g} \quad (40)$$

$$\frac{1}{n} \bar{u} K' = \bar{I} = (1, \dots, 0, \dots, 0)$$

to construct a basis for V_k .

The square multiple correlation of $\bar{u}$ with T is:

$$\underline{h} \in V_n^1 = V_j + V_j^1 \quad (45)$$

$\underline{h}$ can only be written as:

$$\underline{h} = \underline{s} + \underline{k} \quad \underline{s} \in V_j; \underline{k} \in V_j^1 \quad (46)$$

If $\underline{k} = 0$ we have for (23), (24), (25):

$$R_{\underline{h}j}^2 = 1 = R_{\underline{h}(j \oplus k)}^2 = R_{\underline{h}w}^2 \quad (47)$$

For the isomorphism:

$$R_{\underline{h}w}^2 = R_{\underline{h}1}^2 = 1 \quad (48)$$

On the contrary, if $\underline{k} \neq 0$, we construct the vector:

$$\underline{k}^* = \underline{k} (n / \underline{k} \cdot \underline{k})^{1/2} \quad (49)$$

thus:

$$(\underline{k}^* \cdot \underline{k}^*) / n = 1 \quad (50)$$

Like $\underline{k}$ also $\underline{k}^* \in V_j^1$: this one can be used to construct a basis K for V_k with $f+g-1$ other vectors $\underline{b}_j$ such as:

$$\underline{k}j^* = 0; \underline{k} \cdot \underline{b}_j^* = 0; \underline{k}k^* = I_{f+g} \quad (51)$$

As V_j contains $\underline{s}$ and V_k contains $\underline{k}$, we have:

$$\underline{h} \in (V_j + V_k) = V_w \quad (52)$$

and:

$$R_{\underline{h}w}^2 = 1 \quad (53)$$

To conclude for the isomorphism we have:

$$R_{\underline{h}w}^2 = R_{\underline{h}1}^2 = 1 \quad (54)$$

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.

$$R_{\underline{u}t}^2 = \frac{1}{n} \underline{u} \cdot \underline{p}_t \underline{u}' \quad (41)$$

and, for (24):

$$R_{\underline{u}t}^2 = \frac{1}{n} \underline{u} (K'Q_t' + J'S_t^{-1} V_t S_t^{-1} Q_t K + S_t V_t S_t^{-1} J) \underline{u}' \quad (42)$$

For the non correlation between $\underline{u}$ and J and by (40) we have:

$$R_{\underline{u}t}^2 = \underline{q}_{1t}' S_t^{-1} \underline{q}_{1t} \quad (43)$$

where $\underline{q}_{1t}$ is the first row of the matrix Q :

Because the matrix S is positively defined by (43) we have:

$$R_{\underline{u}t}^2 = \underline{q}_{1t}' S_t^{-1} \underline{q}_{1t} > 0 \quad (44)$$

As regards the errors in equation G , a similar demonstration can be performed.

8. THIRD THEOREM

The theorems 1 and 2 demonstrate inconsistent properties of latent variables and errors in equations. If the vectors of the latent variables T and of the errors in equation G are linearly independent, it follows that criteria completely uncorrelated with the observed variables are predictable from the latent variables T and the errors in equations G , while criteria perfectly correlated with observed variables are completely unpredictable from the latent variables. The third theorem demonstrates that the latent variables and errors in equations and in variables can be constructed so as to predict any given criteria $\underline{k}$, including those that are entirely uncorrelated the observed scores J . This happens in spite of the observed score with all the information concerning latent variables and errors.

PROOF

Let V_L and V_W be the spaces generated by the rows of the matrices L and W defined in (24), (25). The mapping from V_L onto V_W is isomorphic, because the matrix Γ in (25) which represents this mapping relative to the basis L and W has an inverse matrix, for the properties of block matrices.

Let $\underline{h}$ be a criterion such that:

Moreover a criterion exists, perfectly correlated with latent variables and errors, and completely uncorrelated with the observed scores.

BIBLIOGRAPHY

Anderson T.W. and Rubin M. (1956). "Statistical inference in factor analysis", in J. Neyman ed., *Proceedings of the third Berkeley Symposium on Mathematical Statistics and Probability*, Vol. V, Berkeley University of California Press, pp. 111-150.

Guttman L. (1955). "The determinacy of factor scores matrices with implications for five others basic problems of common factor theory", *The British Journal of Statistical Psychology*, Vol. 8, pp. 65-81.

Haagen K. (1983). "Il problema dell'identificazione nell'analisi fattoriale e le conseguenze nelle applicazioni empiriche", *Quaderni di Statistica Matematica applicata alle scienze economiche-sociali*, Università di Trento, Vol. 6, pp. 21-28.

Haagen K. (1986). "The indeterminacy Problem in Factor Analysis and its consequences for prediction", *Osterreichische Zeitschrift für Statistik und Informatik*, Vol 15, pp. 197-205.

Heerman E.F. (1964). "The geometry of factorial indeterminacy", *Psychometrika*, Vol. 29, pp. 371-381.

Heerman E.F. (1966). "The algebra of factorial indeterminacy", *Psychometrika*, Vol. 31, pp. 539-543.

Jöreskog K.G. (1973). "A general method for estimating a linear structural equation system", *Biometrika*, Vol. 57, pp. 239-251.

Jöreskog K.G. (1977). "Structural equation models in the Social sciences: specifications, estimation and testing" in P.R. Krishnaiah ed., *Applications of statistics*, Amsterdam, North Holland, pp. 265-287.

Jöreskog K.G. (1978). "Structural analysis of covariance and correlation matrices", *Psychometrika*, Vol. 43, pp. 443-477.

Jöreskog K.G. (1981). "Analysis of covariance structures", *Scandinavian Journal of Statistics*, Vol. 8, pp. 65-92.

Jöreskog K.G. (1982). "The Lisrel Approach to causal model building in the social sciences", in K.G. Jöreskog - H. Wold ed., *Systems under indirect observation: causality, structure prediction*, Amsterdam, North Holland, Cap. 4, pp. 81-99.

Jöreskog K.G. (1982a). *The Lisrel V*. Uppsala University Press.

Jöreskog K.G. and Sörbom D. (1976). "Statistical models and method for analysis of longitudinal data" in D.J. Agner - Golderberger eds., *Latent variables in socio-economic models*, Amsterdam, North Holland, pp. 285-325.

Kestelman H. (1952). "The fundamental equation of factor analysis", *British Journal of Psychology*, Statistical section, pp. 1-6.

Lawley D.N. and Maxwell A.E. (1963). *Factor analysis as a Statistical Method*. Londra, Butterworth.

Ledermann W. (1938). "The orthogonal transformation of a factorial matrix into itself", *Psychometrika*, Vol. 3, pp. 181-187.

Piaggio H.T.W. (1931). "The general factor in Spearman's theory of intelligence", *Nature*, Vol. 127, pp. 56-57.

- Reiersol E. (1950). "On the identifiability of parameters in Thurstone's multiple factor analysis", *Psychometrika*, Vol. 15, pp. 121-149.
- Schönemann P.H. (1971). "The minimum average correlation between equivalent sets of uncorrelated factors". *Psychometrika*, Vol. 35, pp. 21-30.
- Schönemann P.H. and Wang M.M. (1972). "Some new results of factor indeterminacy", *Psychometrika*, Vol. 37, pp. 61-91.
- Schönemann P.H. and Steiger J.H. (1976). "Regression component analysis", *British Journal of Mathematical and Statistical analysis*, Vol. 29, pp. 175-189.
- Schönemann P.H. and Steiger J.H. (1978). "On the validity of indeterminate factor scale", *Bulletin of the Psychonomic society*, Vol. 12, pp. 287-290.
- Schönemann P.H., Haagen K. (1986). On the use of factor scores for prediction. *Universita' degli Studi di Trento, Istituto di Statistica e ricerca operativa*, pro-manuscripto.
- Takeuchi K., Yanay H., Murherjee B.N. (1982). *Multivariate analysis*, New York, Wiley.
- Received by Editorial Board member September, 1987; Revised November, 1987.*
- Recommended by A. M. Kshirsagar.*