

SOMMARIO

Paola Caselli: Rate and Pricing Strategies in a Model of International Duopoly.
Michele Bernasconi: Decision Making under Risk and Uncertainty: A Perspective on the Literature.
Michele Zenke: L'indice Zip) come misura della concentrazione locale.
Pietro Fabbri, Raimondello Orsini: Una rete neurale per studiare l'avversione al rischio.

Riassunti in inglese - Recenti pubblicazioni.

Annals of Human Genetics

CONTENTS

- T. Yasuda, D. Nadano, K. Sawasaki, K. Kishi: Genetic polymorphism of human deoxyribonuclease II (DNase II): low activity levels in urine and leukocytes are due to an autosomal recessive allele.
H.-H. Renger, R. Peacock, A.M. Dunning, P. Talmud, S.E. Humphries: Genetic relationship between the 3' VNTR and diallelic apolipoprotein B gene polymorphisms: Haplotype analysis in individuals of European and South Asian origin.
R.M. Corbo, R. Ingiano, R. Scacchi, A. Bozzi: Kinetic properties of the common electrophoretic variants of human S-adenosylhomocysteine hydrolase (AHCY): the effect of four nucleoside analogue inhibitors.
R. Chakraborty, M. de Andrade, S.P. Deiger, B. Budowle: Apparent heterozygote deficiencies observed in DNA typing data and their implications in forensic applications.
C.M. Lewis, C. Canning: The number of loci needed for ELOD calculations.
A.W.F. Edwards, J.H. Edwards: Metrical Venn diagrams.
I. Barret: Kinship profiles and their errors.

Books received - Review.

UN CONFRONTO FRA UN MODELLO DI ANALISI CAUSALE
CON VARIABILI LATENTI E METODI ALTERNATIVI

Giorgio Vittadini

1. INTRODUZIONE

Particolare attenzione ha incontrato in letteratura negli ultimi anni il problema dell'analisi di relazioni causali di tipo lineare tra due insiemi di variabili latenti (VL). In particolare Jöreskog (1970, 1971, 1973a, 1973b, 1974, 1977, 1978, 1979a, 1979b, 1981, 1982) negli anni '70 ha proposto a tale scopo il modello Lisrel (LM), modello composto da una equazione strutturale e due equazioni di misura. È stato però messo in luce dallo stesso Jöreskog che non esistono condizioni generali necessarie e sufficienti per l'identificazione dei parametri del LM (Jöreskog, 1981). È stato altresì dimostrato che, anche in particolari casi in cui i parametri sono identificati, sussiste il problema dell'indeterminatezza (non unicità) dei punteggi delle VL (Vittadini 1988, 1989). Per evitare tali problemi sono stati proposti metodi alternativi per ricavare le VL del LM: si menzionano in questo lavoro la proposta di Wold del *soft modelling a partial least squares* (PLS) (1980, 1982) e quella di Haagen e Vittadini (1991) della *regression component decomposition* (RCD). Nel primo caso sono ricavate *proxies* per le VL del LM come combinazioni lineari delle variabili osservate in un procedimento iterativo a più stadi; nel secondo caso si ottengono "componenti" in una decomposizione della matrice dei dati. In questo lavoro attraverso una trattazione teorica e alcune esemplificazioni si confrontano il LM e gli altri due metodi. Rispetto ad altri analoghi confronti (Löhmoller, 1989), il criterio utilizzato è quello di verificare se e in che misura le proprietà delle VL sono realmente rispettate dalle variabili ricavate mediante il PLS e l'RCD.

2. IL MODELLO LISREL

Il LM è composto da una equazione strutturale e due equazioni di misura:

$$z_i = Az_i + Pt_i + g_i$$

$$x_i = L_x z_i + d_i$$

$$y_i = L_y t_i + e_i \quad i = 1, 2, \dots, n$$

(1)

dove i vettori z, t, x, y hanno m, u, q, p componenti con $q > m, p > u, n > p + q$.

Si assume che tutte le variabili casuali abbiano media zero e varianza finita. A è matrice triangolare bassa con zero sulla diagonale principale; (t_1, g_1, d_1, e_1) sono identicamente e indipendentemente distribuiti; inoltre, dati $T = (t_1', \dots, t_n')$, $G = (g_1', \dots, g_n')$ (errori nelle equazioni), $D = (d_1', \dots, d_n')$ ed $E = (e_1', \dots, e_n')$ (errori nelle variabili):

ϕ_1 (G, D, E) e T sono indipendenti,
 ϕ_2 (D, E) e G sono indipendenti,
 ϕ_3 D e E sono indipendenti.

Il modello (1) può essere riscritto in forma ridotta:

$$J = LR + H = L_F F, \quad (2)$$

con:

$$L = \begin{pmatrix} L_x(I - A)^{-1}P & L_x(I - A)^{-1} \\ L_y & 0 \end{pmatrix}, \bar{I} = \begin{bmatrix} I \\ O \end{bmatrix}, L_F = [L, \bar{I}],$$

$$R = \begin{bmatrix} T \\ G \end{bmatrix}, H = \begin{bmatrix} D \\ E \end{bmatrix}, F = \begin{bmatrix} R \\ H \end{bmatrix}.$$

Dati le (ϕ_1) , (ϕ_2) , (ϕ_3) , L , $L_x(I - A)^{-1}P$, $L_x(I - A)^{-1}$, L_y sono rispettivamente matrici di regressione di J su R , X su T , X su G , Y su T .

La matrice di varianze-covarianze del modello è quindi:

$$\Sigma_J = L \Sigma_R L' + \Sigma_H = L_F \Sigma_F L_F', \quad (3)$$

con:

$$\Sigma_R = \begin{bmatrix} \Sigma_T & 0 \\ 0 & \Sigma_G \end{bmatrix}, \Sigma_H = \begin{bmatrix} \Sigma_D & 0 \\ 0 & \Sigma_E \end{bmatrix}.$$

2.1. Non identificabilità dei parametri e indeterminanza del Modello Lisrel

2.1.1. Identificabilità

In generale si assume che la distribuzione delle variabili osservate sia sufficientemente ben descritta dai momenti primo e secondo (in particolare ciò vale nell'ipotesi che la distribuzione sia normale multivariata), cosicché le informazioni aggiuntive riguardanti i parametri da stimare contenute nei momenti di ordine superiore possono essere ignorate. Per affrontare il problema dell'identificabilità dei parametri del LM quando le variabili osservate siano centrate, è sufficiente verificare se, data una matrice di varianze-covarianze delle variabili osservate Σ_J , il vettore θ dei parametri e delle varianze-covarianze delle variabili latenti è determinato univocamente. Data la matrice di varianze-covarianze Σ_J , esistono $1/2(p+q)(p+q+1)$ equazioni $\sigma_{ij} = \sigma_{ij}(\theta)$ ($i \leq j$) in n_θ sconosciuti parametri (n_θ dipendente dalla assunzione sull'incorrelazione tra

singole variabili latenti ed errori). Condizione necessaria per l'identificabilità è che:

$$n_\theta \leq \frac{1}{2}(p+q)(p+q+1). \quad (4)$$

Per identificare i parametri, Jöreskog (1981) propone di porre un numero opportuno di parametri pari a zero od uno e di vincolarne altri ad essere uguali l'un l'altro. Tuttavia "soluzioni esplicite per tutti gli θ , raramente esistono" (Jöreskog, 1981, p. 73) in quanto "non esistono condizioni, necessarie e sufficienti, generali e praticamente utili per l'identificabilità di ogni parametro del modello" (Jöreskog, 1981, p. 89). Perciò Jöreskog suggerisce che "il problema dell'identificazione sia affrontato caso per caso sulla base delle singole equazioni".

2.1.2. Indeterminanza

Si osservi che sotto condizioni generali nel LM

$$\text{rango}(\Sigma_J) = \text{rango}(L_F \Sigma_F L_F') = (q+p) < \text{rango}(\Sigma_F) = (q+p+m+u), \quad (5)$$

vale a dire che le $q+p$ variabili osservate sono "spiegate" da $q+p+m+u$ vettori di VL ed errori tra loro linearmente indipendenti. Proprio la differenza di rango tra $\Sigma_J = (L_F \Sigma_F L_F')$ e Σ_F comporta che in generale il numero di vettori linearmente indipendenti in F ecceda il numero di vettori linearmente indipendenti in J . Perciò non è possibile determinare una soluzione unica in F per il modello, anche quando sia identificata per una particolare matrice di varianze-covarianze Σ_F .

La non unicità o "indeterminanza" delle VL del LM, le cui conseguenze sulla capacità predittiva e classificatoria, e in alcuni casi sul significato stesso del modello, sono assai rilevanti (Schönemann e Haagen, 1987; Haagen, 1991), è messa in luce dal seguente lemma:

$$R = \Sigma_R L' \Sigma_J^{-1} J + V \quad (6)$$

e

$$(\Sigma_R - \Sigma_R L' \Sigma_J^{-1} L \Sigma_R) \neq 0, \quad (7)$$

con V parte arbitraria delle VL R ortogonali a J , ineliminabile con un numero finito di variabili osservate e VL (Haagen, 1991).

2.2. Esempio

Si considera un esempio proposto da Jöreskog (1982) inerente alla "stabilità dell'alienazione" dove compaiono le seguenti variabili:

x_1, x_2 = mancanza di motivazioni '67, '71;

x_3, x_4 = povertà '67, '71;

y_1 = livello di educazione;

y_2 = indice socio-economico di Duncan;

t = status socio-economico;

z_1, z_2 = "alienazione" '67, '71.

Il modello ha la seguente struttura:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ l_1 & 0 \\ 0 & 1 \\ 0 & l_2 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} d_1 \\ d_2 \\ d_3 \\ d_4 \end{bmatrix}$$

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 \\ l_3 \end{bmatrix} t + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ a & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} t + \begin{bmatrix} g_1 \\ g_2 \end{bmatrix},$$

con x_i, y variabili osservate, z, t variabili latenti, g errori nelle equazioni, e e d errori nelle variabili. Le scale per z_1, z_2, t sono state scelte identiche rispettivamente a x_1, x_3 e y_1 (ponendo i rispettivi parametri pari a 1). Oltre alle assunzioni usuali per il LM, si assume che gli errori e_i, g_i, d_i siano incorrelati.

Il modello in forma ridotta ha sei variabili osservate con ventuno varianze-covarianze di variabili osservate e quindici parametri e varianze di VL ($a, 2b_i, 3l_i, \sigma_e, 2\sigma_g, 4\sigma_d, 2\sigma_t$); si dimostra che è identificato. La matrice di varianze-covarianze campionarie delle variabili osservate Σ_j è:

TABELLA 1
Matrice di varianze-covarianze tra variabili osservate

	x_1	x_2	x_3	x_4	y_1	y_2
x_1	11.834					
x_2	6.946	9.363				
x_3	6.819	5.091	12.532			
x_4	4.783	5.028	7.454	9.985		
y_1	-3.839	-3.889	-3.841	-3.625	9.61	
y_2	-21.900	-18.829	-21.746	-18.771	35.52	450.289

Risolvendo mediante il metodo della massima verosimiglianza si ottengono i seguenti valori per i parametri:

TABELLA 2
Parametri e varianze delle VL del LM

$$l_y = \begin{bmatrix} 1 \\ 5.329 \end{bmatrix}; A = \begin{bmatrix} 0 & 0 \\ 0.705 & 0 \end{bmatrix}; P = \begin{bmatrix} -0.614 \\ -0.714 \end{bmatrix}; L_x = \begin{bmatrix} 1 & 0 \\ 0.889 & 0 \\ 0 & 1 \\ 0 & 0.849 \end{bmatrix}$$

$$\sigma_{FF} = [5.307 \quad 3.742 \quad 6.610 \quad 4.015 \quad 3.192 \quad 3.701 \quad 3.625 \quad 2.944 \quad 260.982]; \quad \sigma_{FQ} = 0 \quad (F \neq 0)$$

L'indeterminatezza può essere verificata dal fatto che il modello ammette soluzioni non univoche. Ricavata da Σ_j una matrice di variabili riprodotte J^* (tabella 3) tale che $\frac{1}{n} J^+ J^{*'} = \Sigma_j$, si osserva che ad esempio esistono due matrici di soluzione R^0 ed $\bar{R}$ tali che $R^0 R^{0'} = \bar{R} \bar{R}' = \Sigma_j$, le cui corrispondenti correlazioni sono molto lontane da uno (tabella 4).

TABELLA 3
Variabili riprodotte J^*

1.012	.054	5.245	4.485	-3.331	-2.518	-5.520
2.195	.881	1.026	6.167	-7.010	-.522	-2.906
-2.859	-5.095	3.726	3.405	-5.312	-2.865	-4.112
-4.547	-1.954	3.222	5.763	-3.434	-4.410	.382
-3.914	.384	-2.238	-4.977	-2.176	3.187	4.282
-28.443	28.719	-15.987	-26.897	-24.591	-9.407	46.809
-5.468	5.566	.399	2.573	-.728	-4.569	-4.639
-.936	3.995	2.113	1.866	2.777	1.070	-4.076
2.132	1.126	5.544	-.167	-5.102	-.960	-.358
2.982	6.578	2.125	-1.338	-1.631	-.581	-1.187
-4.224	-2.130	-4.942	3.235	.732	2.648	2.890
6.953	-17.896	-28.333	.867	-1.724	-22.272	26.594
-2.934	1.327	4.584	1.581	1.378	-.299	.340
-.915	1.056	4.126	-4.279	-.688	-2.807	-.328
-.327	.246	4.115	-.524	3.115	-3.220	.120
3.274	-1.397	1.652	-.466	1.723	-3.028	-1.081
-.125	1.159	-4.889	1.293	-2.199	.822	3.528
-18.102	22.802	13.236	-9.587	-13.151	17.993	27.650
5.293	-2.098	.999	-2.732			
2.791	-3.237	.433	-2.773			
9.047	.223	.276	-2.173			
3.141	-1.288	.395	-6.058			
1.334	5.406	1.617	-.703			
-10.527	13.125	14.739	7.429			

TABELLA 4
Correlazioni tra due soluzioni R^0 e $\bar{R}$ del LM

$$\text{cor}(R^0, \bar{R}) = 10.586, 0.439, 0.281, 0.728, 0.447, 0.443, 0.505, 0.416, 0.478]$$

3. MODELLO "SOFT MODELLING" A "PARTIAL LEAST SQUARES"

Wold (1980, 1982) introduce in un'ottica più generale che non quella di proporre una alternativa al Lisrel, un modello denominato *soft modelling*, utilizzabile quando le conoscenze sono scarse e non sono applicabili assunzioni stringenti a riguardo della distribuzione delle variabili. Nel caso si abbiano due insiemi di VL, ognuno composto da una o più VL, il modello di *path analysis*,

come nel LM, ha la seguente struttura (Apel e Wold, 1982; Hui e Wold, 1982):

$$Z = AZ + PT + G, \quad (8)$$

con A matrice triangolare bassa come nel LM; la s -esima VL z_s è funzione lineare di tutte le VL z_l tali che $l < s$ e di tutte le variabili t_k , a meno di un errore g_s :

$$z_s = \sum_{l=1}^{s-1} a_{sl} z_l + \sum_k p_{sk} t_k + g_s \quad (s = 1, 2, \dots, m). \quad (8a)$$

Nei modelli di misura del LM le variabili osservate $X_k(Y)$, i cui coefficienti di regressione rispetto alle VL $Z_k(T)$ non sono posti a priori uguali a zero, costituiscono un insieme di indicatori multipli per tali VL (ogni variabile osservata può appartenere contemporaneamente a più di un insieme $X_k(Y)$):

$$b^X = b^L X + b^D \quad (b = 1, 2, \dots, n_b) \quad (9a)$$

$$k^Y = k^L Y + k^E \quad (k = 1, 2, \dots, n_k), \quad (9b)$$

con:

$$Z = [z_1', \dots, z_m', \dots, z_m']', \quad T = [t_1', \dots, t_k', \dots, t_k']',$$

$$D = [d_1', \dots, d_{n_b}', \dots, d_{n_b}']', \quad E = [e_1', \dots, e_{n_k}', \dots, e_{n_k}']',$$

$$L_x = [l_1^{L_x}, \dots, l_{n_b}^{L_x}], \quad l^{L_x} = [l_1^{L_x}, 0], \quad l^{L_x} = [0, l^{L_x}, 0], \quad n_b^{L_x} = [0, n_b^{L_x}, 1];$$

$$L_y = [l_1^{L_y}, \dots, l_{n_k}^{L_y}], \quad l^{L_y} = [l_1^{L_y}, 0], \quad l^{L_y} = [0, l^{L_y}, 0], \quad n_k^{L_y} = [0, n_k^{L_y}, 1].$$

Tutti gli errori E, D, G si assumono tra loro incorrelati e incorrelati con le VL Z, T (eccetto gli z_s e g_s appartenenti alla medesima equazione (8a)):

$$(w_1) \text{cor}(T, G) = \text{cor}(T, E) = \text{cor}(T, D) = \text{cor}(G, E) = \text{cor}(G, D) = \text{cor}(E, D) = 0,$$

$$\text{cor}(Z, D) = \text{cor}(Z, E) = \text{cor}(z_l, g_s; l \neq s) = 0, \quad (S_G, S_E, S_D) = \text{diag}.$$

Ciò premesso per quanto concerne la specificazione del *soft modelling*, Wold propone di ricavare, come *proxies* per le VL z_h, t_k , combinazioni lineari delle variabili osservate $\hat{z}_h, \hat{t}_k$, mediante un metodo iterativo di "stima" basato sui minimi quadrati denominato *partial least squares* (PLS). Tali *proxies* non coincidono con le VL ($\text{cor}(t_k, \hat{t}_k), \text{cor}(z_h, \hat{z}_h) < 1$) ma ne costituiscono approssimazioni che non godono necessariamente di tutte le proprietà delle VL medesime.

3.1. Metodo per ricavare VL e parametri

3.1.1. Primo stadio PLS: Metodo per ricavare le "proxies"

Le singole *proxies* vengono "stimate" (approssimate) come combinazioni lineari delle variabili osservate loro indicatori. Si consideri, per semplicità, il caso in cui, essendo ogni sottinsieme di variabili osservate indicatore di una sola VL, si ha in (9a)-(9b) $b^Z \hat{z}_h, k^T \hat{t}_k = \hat{t}_k$. Si ottiene:

$$\hat{z}_h = \sum_f b^Z w_f b^X x_f, \quad \hat{t}_k = \sum_j k^T w_j k^Y y_j, \quad (10)$$

con pesi $b^Z w_f, k^T w_j$ ricavati (modo A) attraverso regressioni multiple del tipo:

$$\hat{z}_h = \sum_f b^Z w_f b^X x_f + b^Z \gamma, \quad \hat{t}_k = \sum_j k^T w_j k^Y y_j + k^T \gamma, \quad (11)$$

o (modo B) attraverso regressioni semplici del tipo:

$$b^X x_f = b^Z w_f \hat{z}_h + b^Z \delta_f \quad (f = 1, 2, \dots, n_f),$$

$$k^Y y_j = k^T w_j \hat{t}_k + k^T \delta_j \quad (j = 1, 2, \dots, n_j). \quad (12)$$

In entrambi i casi, le suddette regressioni (11)-(12) sono all'interno del seguente procedimento iterativo a più passi:

Passi 1 e 2:

Al passo 2 ogni *proxy* è calcolata come somma pesata delle *proxies* ad esse adiacenti (direttamente connesse con essa nel diagramma *path*):

$$\hat{z}_h = \sum_l v_{hl} \hat{z}_l + \sum_q v_{hq} \hat{t}_q, \quad \hat{t}_k = \sum_l v_{kl} \hat{z}_l + \sum_q v_{kq} \hat{t}_q. \quad (13)$$

I pesi $v_{hl}, v_{hk}, v_{kl}, v_{kq}$ sono ricavati al passo 1 comunemente secondo uno dei seguenti metodi (o un loro mix) (Noonan e Wold, 1982, pp. 75-79):

ω_1 coefficienti di regressione multipla (o semplice) di ogni $\hat{z}_h, \hat{t}_k$ sulle sue "cause" nella *path analysis* (8);

$$\omega_2) \quad v_{hl} = \text{cor}(\hat{z}_h, \hat{z}_l), \quad v_{hq} = \text{cor}(\hat{z}_h, \hat{t}_q), \quad v_{kl} = \text{cor}(\hat{t}_k, \hat{z}_l), \quad v_{kq} = \text{cor}(\hat{t}_k, \hat{t}_q);$$

$$\omega_3) \quad v_{hl}, v_{hq}, v_{kl}, v_{kq} = \pm 1 \text{ a seconda del segno dei coefficienti di correlazione indicati in } \omega_2).$$

Passo 3:

Al passo 3 si effettuano le regressioni multiple (11) delle $\hat{z}_h$ e $\hat{t}_k$ ottenute al passo 2 sulle variabili osservate loro indicatori o le regressioni semplici (12) delle variabili osservate indicatori sulle $\hat{z}_h, \hat{t}_k$ ottenute al passo 2.

Passo 4:

Al passo 4 si ricavano nuove *proxies* attraverso i pesi $b^Z w_f, k^T w_j$ ottenuti al passo 3.

$$\hat{z}_k = \pi_b \sum_f b^Z w_f b^X x_f, \quad \hat{t}_k = \pi_k \sum_j k^T w_j k^Y y_j, \quad (14)$$

con π_b, π_k scalari tali che $\text{var}(\hat{z}_b) = \text{var}(\hat{t}_k) = 1$.

Le iterazioni si interrompono quando i cambiamenti di valore di $b^Z w_f, k^T w_j$ sono irrilevanti.

3.1.2. Secondo stadio PLS: metodo per ricavare i parametri

Nel secondo stadio del PLS sono ricavate le matrici di parametri $\hat{L}_x, \hat{L}_y$ e gli errori b^E e k^E attraverso regressioni semplici ai minimi quadrati ordinari (OLS) di ogni variabile osservata $b^X x_f (k^Y y_j)$ sulle VL $\hat{z}_b (\hat{t}_k)$. Le matrici di parametri $\hat{A}$ e $\hat{P}$ e gli errori nel modello di *path analysis* (8) $\hat{G}$ sono ricavati attraverso

successive regressioni OLS delle variabili dipendenti sulle variabili indipendenti in (8a).

3.1.3. Caso di più VL per ogni sottoinsieme di variabili osservate

Nel caso in cui ogni sottoinsieme di variabili osservate è indicatore di più VL, si effettua il procedimento descritto nel paragrafo 3.1.1 per ricavare le prime VL degli insiemi $\hat{Z}_k(\hat{T})$, $\hat{Z}_h(\hat{I}_k)$. Effettuate poi le regressioni semplici OLS di ogni variabile osservata y_j sulle VL $\hat{Z}_h(\hat{I}_k)$, nelle formule (10)-(14) si utilizzano i residui di tali regressioni in luogo delle variabili x_j per ottenere le "stime" delle seconde VL $\hat{Z}_h(\hat{I}_k)$. Il procedimento continua finché non sono stimate le ultime VL di cui i sottoinsiemi sono indicatori. A questo punto si ricavano gli errori $G, \hat{E}, \hat{D}$ come residui in (8), (9a), (9b).

3.2. Discussione

3.2.1. PLS: metodo per ricavare combinazioni lineari delle variabili osservate, non fattori causali

Come si evince dalla formula (14), le variabili le *proxies* delle VL nel PLS sono ricavate come combinazioni lineari delle variabili osservate

$$\text{range } \Sigma_J = \text{range}(\hat{L}_F \hat{\Sigma}_F \hat{L}_F') = (q + p) = \text{range } \Sigma_F. \quad (15)$$

Comunque siano definite le VL nel *soft modelling*, le loro "approssimazioni" non possono essere intese come "fattori causali" non osservabili che determinano variazioni e covariazioni delle variabili osservate. Ne consegue che i modelli di misura (9), sostituendo alle VL le loro *proxies*, si riducono a metodi descrittivi di analisi in cui le $p + q$ variabili osservate sono spiegate dalle loro trasformate lineari; la *path analysis* (8) diviene un modello di analisi causale tra due insiemi di combinazioni lineari delle variabili osservate, l'uno dipendente e l'altro indipendente.

3.2.2. Indeterminatezza

Quanto ricordato nel paragrafo precedente a riguardo dell'uguaglianza del rango di $\hat{\Sigma}_F$ e Σ_J comporta il superamento del problema dell'indeterminatezza delle VL, riscontrato nel LM. Infatti a differenza di quanto accade in (7), si osserva che:

$$\hat{\Sigma}_{\hat{R}} - \hat{\Sigma}_{\hat{R}} \hat{C}_{\hat{R}} \Sigma_J^{-1} \hat{C}_{\hat{R}}' \Sigma_{\hat{R}} = 0, \quad (16)$$

con $C_{\hat{R}}$ matrice di regressione di J su $\hat{R}$.

3.2.3. Identificabilità

Definendo le $\hat{Z}_h$ e le $\hat{I}_k$ come trasformazioni lineari delle variabili osservate attraverso il procedimento iterativo sopra descritto, viene superato anche il

problema della identificabilità dei parametri tipico del LM. Infatti nelle equazioni (9) i parametri sono ricavati attraverso regressioni semplici OLS delle variabili osservate rispetto alle $\hat{Z}_h, \hat{I}_k$ loro indicatori; nella *path analysis* (8) i parametri sono ricavati attraverso successive regressioni multiple OLS delle variabili dipendenti sulle variabili indipendenti in (8a). L'unico ulteriore possibile motivo di non identificabilità del modello, dovuto alla scala delle *proxies*, è poi eliminato imponendo la standardizzazione delle *proxies* medesime nel passo 4 del primo stadio del PLS.

3.3. Compatibilità del PLS con il LM

Per quanto detto, il PLS ha un ben preciso significato e una sua coerenza rispetto alle assunzioni del *soft modelling* (8)-(9): tuttavia appare inadeguato quando si voglia usare come metodo alternativo per ricavare le VL del LM. Infatti, le approssimazioni delle VL ricavate mediante il PLS non soddisfano numerose proprietà richieste alle VL del LM.

3.3.1. Definizione delle "proxies"

Il primo problema di ordine generale nasce dal modo con cui si definiscono le *proxies* delle VL nei passi 1 e 2 del primo stadio del procedimento iterativo di stima PLS. Pur prescindendo dal fatto che, come segnala lo stesso Wold (Noonan e Wold, 1982), non esistono modi oggettivi per scegliere tra i diversi modi per assegnare i pesi $v_{h1}, v_{h2}, \dots, v_{hp}$, appare comunque discutibile definire le variabili "esogene" $\hat{I}_k$ in (8a) come combinazioni lineari delle variabili $\hat{Z}_l$ e $\hat{I}_h$ adiacenti, di cui sono in realtà cause.

3.3.2. Incorelazione fra $\hat{T}, \hat{G}, \hat{E}, \hat{D}$

Si ricorda che, mentre le soluzioni del LM sono ricavate a partire dalla matrice di varianze-covarianze del modello ridotto (3), attraverso il PLS si ricavano dapprima gli *scores* delle *proxies* delle variabili Z e T come combinazioni lineari delle variabili osservate; successivamente si ottengono parametri ed errori, risolvendo con il metodo OLS le equazioni del modello strutturale (8a) e dei modelli di misura (9a)-(9b).

Ciò premesso, si definiscono in coerenza con (14) le combinazioni lineari $\hat{b}\hat{Z} = {}_bW_bX$ ($b = 1, 2, \dots, n_b$) e $\hat{k}\hat{T} = {}_kW_kY$ ($k = 1, 2, \dots, n_k$); i proiettori ortogonali su $S(\hat{Z})$ e $S(\hat{T})$ spazi generati dalle righe di $\hat{b}\hat{Z}$ e $\hat{k}\hat{T}$

$$P_{\hat{b}\hat{Z}} = \hat{b}\hat{Z}(\hat{b}\hat{Z}\hat{b}\hat{Z}')^{-1} \hat{b}\hat{Z} = {}_bX'{}_bW_b'({}_bW_bX{}_bX'{}_bW_b')^{-1} {}_bW_bX,$$

$$P_{\hat{k}\hat{T}} = \hat{k}\hat{T}(\hat{k}\hat{T}\hat{k}\hat{T}')^{-1} \hat{k}\hat{T} = {}_kY'{}_kW_k'({}_kW_kY{}_kY'{}_kW_k')^{-1} {}_kW_kY;$$

i proiettori sui complementi ortogonali a $S(\hat{Z})$ e $S(\hat{T})$ $Q_{\hat{b}\hat{Z}} = I - P_{\hat{b}\hat{Z}}$ e $Q_{\hat{k}\hat{T}} = I - P_{\hat{k}\hat{T}}$. Si ottiene quindi:

$$\hat{Z}_s = \Sigma_J \hat{\Sigma}_J^{-1} \hat{P}_{\hat{k}\hat{T}} \hat{Z}_k + \hat{G}_s = \hat{Z}_s P_{\hat{k}\hat{T}} + \hat{Z}_s Q_{\hat{k}\hat{T}} \quad (s = 1, 2, \dots, m), \quad (17)$$

$${}_bX = {}_b\hat{L}_{xb}\hat{Z} + {}_b\hat{D} = {}_bXP_{\hat{Z}} + {}_bXQ_{\hat{Z}} \quad (18)$$

$${}_kY = {}_k\hat{L}_{yk}\hat{T} + {}_k\hat{E} = {}_kYP_{\hat{T}} + {}_kYQ_{\hat{T}} \quad (19)$$

Dalla (17) e (18) consegue inoltre che:

$$\text{cor}(\hat{T}, \hat{G}) = \text{cor}({}_k\hat{T}, {}_k\hat{E}) = 0 \quad (20)$$

Per il resto, per come sono definite le $\hat{i}_k$ in (14), le $\hat{g}_s$ in (17), le $\hat{D}$ in (18), le ${}_k\hat{E}$ in (19), si ha sotto condizioni generali:

$$\text{cor}({}_b\hat{T}, {}_b\hat{E}) \neq 0 \quad (b \neq k, \quad b = 1, 2, \dots, n_b, \quad k = 1, 2, \dots, n_k), \quad (21)$$

$$\text{cor}(\hat{T}, \hat{D}) \neq 0, \quad (22)$$

$$\text{cor}(\hat{G}, \hat{D}) \neq 0, \quad (23)$$

$$\text{cor}(\hat{E}, \hat{D}) \neq 0, \quad (24)$$

da cui si deduce che nessuna delle proprietà (ϕ_1) , (ϕ_2) , (ϕ_3) delle VL del LM e non tutte le proprietà (ψ_1) delle VL del *soft modelling* sono rispettate dalle *proxies* ricavate mediante il PLS.

3.3.3. Matrici dei parametri

Una terza questione riguarda la natura delle matrici dei parametri alla fine del procedimento di stima. Ogni sottomatrice di ${}_b\hat{L}_x({}_k\hat{L}_y, {}_b\hat{L}_x({}_k\hat{L}_y))$ è matrice dei coefficienti di regressione di ${}_bX$ su $\hat{z}_kY$ su $\hat{i}_k$. Tuttavia, per quanto detto a proposito della non incorrelazione di variabili ed errori relativi a diversi insiemi sotto condizioni generali $\hat{L}_x, \hat{L}_y, \hat{L}_x(I - \hat{A})^{-1}\hat{P}, \hat{L}_x(I - \hat{A})^{-1}, \hat{L}$ non sono matrici di regressione di X su $\hat{Z}$, Y su $\hat{T}$, X su $\hat{T}$, X su $\hat{G}$, J su $\hat{R}$, in contrasto con quanto avviene nel LM.

3.4. Esempio

Si applica il PLS all'esempio proposto per il LM utilizzando al passo 3 del procedimento di stima delle VL sia la regressione semplice che la regressione multipla e ricavando ai passi 1 e 2 i pesi secondo i tre diversi metodi $\omega_1, \omega_2, \omega_3$.

Le variabili $\hat{z}_1, \hat{z}_2$ calcolate con i sei metodi descritti sono pressoché perfettamente correlate; le variabili $\hat{i}$ calcolate mediante regressione multipla hanno una correlazione che si aggira sullo 0.75 con le variabili $\hat{i}$ calcolate mediante regressione semplice.

Si prende ora in considerazione il procedimento di stima con regressione multipla e pesi ricavati attraverso un mix dei metodi ω_1 e ω_2 (MIX MUL). Si ottengono le combinazioni lineari riportate nella tabella 5 e le matrici dei parametri riportate nella tabella 6.

TABELLA 5

Pesi delle combinazioni lineari ottenuti mediante PLS (MIX MUL)

$$\begin{aligned} \hat{z}_1 &= 0.17972x_1 + 0.15608x_2 \\ \hat{z}_2 &= 0.1735x_1 + 0.15144x_2 \\ \hat{i} &= 0.00435y_1 + 0.04678y_2 \end{aligned}$$

TABELLA 6

Parametri ottenuti attraverso PLS (MIX MUL)

$$\begin{aligned} \hat{L}_x &= \begin{bmatrix} 3.210 & 0 \\ 2.710 & 0 \\ 0 & 3.309 \\ 0 & 2.812 \end{bmatrix}; & \hat{L}_x(I - \hat{A})P &= \begin{bmatrix} -1.050 \\ -0.886 \\ -1.042 \\ -0.885 \end{bmatrix}; & \hat{L}_y &= \begin{bmatrix} 1.704 \\ 21.218 \end{bmatrix} \\ \hat{A} &= \begin{bmatrix} 0 & 0 \\ 0.556 & 0 \end{bmatrix}; & \hat{L}_x(I - \hat{A})^{-1} &= \begin{bmatrix} 3.210 & 0 \\ 2.710 & 0 \\ 1.860 & 3.309 \\ 1.517 & 2.812 \end{bmatrix}; & \hat{P} &= \begin{bmatrix} -0.327 \\ -0.133 \end{bmatrix} \end{aligned}$$

Calcolando i coefficienti di correlazione tra le diverse variabili ed errori, si ottiene la matrice riportata in tabella 7, dalla quale si verifica che, come ricordato in 3.3.2, non sono rispettate dalle *proxies* del PLS le proprietà (ϕ_1) , (ϕ_2) , (ϕ_3) e (ψ_1) .

TABELLA 7

Matrici di correlazione ottenute attraverso PLS (MIX MUL)

i	$\hat{z}_1$	$\hat{z}_2$	$\hat{i}_1$	$\hat{i}_2$	$\hat{d}_1$	$\hat{d}_2$	$\hat{d}_3$	$\hat{d}_4$	$\hat{d}_5$
$\hat{z}_1$	0								
$\hat{z}_2$	0	0							
$\hat{d}_1$	0.10	0.03	-0.017						
$\hat{d}_2$	-0.10	-0.03	0.016	-1					
$\hat{d}_3$	0.06	0.033	-0.019	0.252	-0.252				
$\hat{d}_4$	-0.06	-0.033	0.019	-0.252	0.252	-1			
$\hat{e}_1$	0	-0.302	-0.131	0.107	-0.107	0.052	-0.052		
$\hat{e}_2$	0	0.298	0.130	-0.094	0.095	-0.041	0.041	-1	

TABELLA 8

Matrici di regressione ottenute attraverso PLS (MIX MUL)

$\frac{1}{n} \hat{X} \hat{X}'^{-1}$	$\frac{1}{n} \hat{X} \hat{X}'^{-1}$
$\begin{bmatrix} 3.227 & -0.028 \\ 2.691 & 0.031 \\ 0.056 & 3.275 \\ -0.065 & 2.851 \end{bmatrix}$	$\begin{bmatrix} -1.041 \\ -0.897 \\ -1.034 \\ -0.894 \end{bmatrix}$
$\frac{1}{n} \hat{Y} \hat{Y}'^{-1}$	$\frac{1}{n} \hat{X} \hat{X}'^{-1} \hat{P}$
$\begin{bmatrix} 1.704 \\ 21.218 \end{bmatrix} = \hat{L}_y$	

4. "REGRESSION COMPONENT DECOMPOSITION"

Sulla scia di una proposta di Schönemann e Steiger (1976) valida per il modello fattoriale, Haagen e Vittadini (1991) hanno suggerito come alternativa al LM la *regression component decomposition* (RCD) della matrice dei dati osservati in componenti che hanno proprietà analoghe a quelle del LM.

Utilizzando la stessa notazione utilizzata nel paragrafo 3, e ricordando che P_{TUY} è il proiettore ortogonale sulla somma lineare di $S(T)$ e $S(Y)$ tale che:

$$P_{TUY} = P_{Y|T} + P_T,$$

si ha la seguente decomposizione:

$$S(T, Y, G, X) = S(T) \oplus S(YQ_T) \oplus S(GQ_{TUY}) \oplus S(XQ_{TUYUG}), \quad (25)$$

e inoltre:

$$J = JP_T + JP_{Y|T} + JP_{G|TUY} + JP_{X|TUYUG}, \quad (26)$$

che è riasprimibile nel seguente modo:

$$J = C_{JT}T^* + C_{JG}G^* + (E^*, D^*), \quad (27)$$

con $C'_{JT} = (C'_{XT}, C'_{YT})'$ e $C_{JG} = (C'_{XG}, 0)'$ matrici di regressione rispettivamente di J sulle componenti T^* e di J sulle "componenti" $G^* = \tilde{G}Q_{TUY}$; C'_{XT} , C'_{XG} e C'_{YT} matrici di regressione rispettivamente di X su T^* , X su G^* , Y su T^* (i simboli T^* , G^* , E^* , D^* sono utilizzati per distinguere le "componenti" delle analoghe grandezze nel LM e nel PLS).

4.1. Metodo per ricavare VL e parametri

Si calcolano innanzitutto le componenti T^* . Per la decomposizione (25)-(27) T^* deve essere una base di $S(T)$ tale che:

$$JP_T = (C'_{XT}, C'_{YT})'T^* \quad (28)$$

T^* può essere quindi ricavata secondo un qualunque metodo di estrazione dei *factor loadings* e dei *factor scores* come combinazione lineare delle righe di J :

$$T^* = M^*J, \text{ (con } \Sigma_T^{-1} = M^* \Sigma_J M^*) \quad (29)$$

matrice di varianze-covarianze positiva definita),

qualora su J si effettui la seguente decomposizione:

$$J = C_{JT}T^* + (J - C_{JT}T^*), \quad (30)$$

chiamata *regression component decomposition* (RCD).

Essendo per la (30):

$$C_{JT} = \Sigma_J M^* \Sigma_T^{-1}, \quad (31)$$

si ha per M^* , attraverso la (29) e la (31):

$$M^* = \Sigma_T^{-1} C'_{JT} \Sigma_J^{-1} = (C'_{JT} \Sigma_J^{-1} C_{JT})^{-1} C'_{JT} \Sigma_J^{-1}, \quad (32)$$

e quindi per T^* :

$$T^* = (C'_{JT} \Sigma_J^{-1} C_{JT})^{-1} C'_{JT} \Sigma_J^{-1} J. \quad (33)$$

Applicando la RCD a $(J^* = J - JP_T - JP_{Y|T})$ si ottiene poi la combinazione lineare:

$$G^* = FJ^*, \quad (34)$$

con:

$$F = (C'_{JG} \Sigma_J^{-1} C_{JG})^{-1} C'_{JG} \Sigma_J^{-1}. \quad (35)$$

Dalla (26) si ricavano quindi gli errori D^* ed E^* :

$$E^* = JP_{Y|T}; \quad D^* = JP_{X|TUYUG}. \quad (36)$$

A questo punto si possono ottenere le matrici dei parametri $C_{XT} = L_X^*(I - A^*)^{-1}P^*$, $C_{XG} = L_X^*(I - A^*)^{-1}$, $C_{YT} = L_Y^*$ come matrici di regressione rispettivamente di X su T^* , X su G^* , Y su T^* ($C_{YG} = 0$ per come sono definite le G^*). Si ottiene quindi, mediante $C_{XT} = C_{XG}P^*$, la matrice P^* come:

$$P^* = (C'_{XG}C_{XG})^{-1}C'_{XG}C_{XT}. \quad (37)$$

Infine, dal modello di *path analysis* (8) si ottiene:

$$(I - A^*)ZZ'(I - A^*)' = (P^*T^* + G^*)(P^*T^* + G^*)'. \quad (38)$$

Ponendo $ZZ' = I$, si ottiene $(I - A^*)$ attraverso una fattorizzazione di Cholesky.

Infine, mediante C_{XG}^* si ottiene $L_X^* = C_{XG}^*(I - A^*)$.

TABELLA 9

Variabili ed errori ϵ^* , G^* , D^* , E^* ottenuti mediante RCD

ϵ^*	ϵ^*	G^*	D^*	E^*	ϵ^*	G^*	D^*	E^*	ϵ^*	G^*	D^*	E^*
-0.031	-0.576	1.139	1.822	-1.301	-0.853	-1.277	0.026					
1.468	1.075	-0.108	-0.344	-0.289	-1.013	0.001	-0.033					
1.244	-0.323	0.540	-0.823	-0.316	1.523	-0.760	0.040					
-1.047												
1.668	1.210	-1.100	0.811	-1.490	1.118	-0.060						
1.035	0.856	0.934	-0.367	0.893	0.002	-1.157						
-0.565	0.030	-0.231	0.783	2.137	1.169	-0.941						
-2.512	0.521	-1.130	0.845	0.182	-1.915	-0.955						
-0.380	0.427	0.286	-1.804	-1.110	0.371	0.004						
-1.388	0.519	0.480	1.313	0.002	1.129	0.295						
-0.913	-0.832	0.005	0.146									
-0.036	-0.244	0.228	0.466									
0	0	0	0	0	0.0002	0						
0	0	0	-1.221	-1.065	1.182	-1.792						
1.226	-1.236	-0.489	1.228	1.071	-1.189	1.802						
-1.233	1.263	0.492		0	0	0						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
0.126	-3.170	1.757	0.449	-1.126	-0.085	0.553						
-0.127	3.188	-1.767	-0.452	1.133	0.086	-0.556						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
-1.607	0.796	1.047	-0.753	0.322	-0.218	0.497						
1.617	-0.801	-1.053	0.757	-0.324	0.219	-0.500						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
0	0	0	0	0	0	0						
2.200	0.587	-0.177	2.216									
-2.213	-0.590	0.178	-2.229									
0	0	0	0									
0	0	0	0									
-0.934	0.105	-0.032	-0.291	-1.162	0.051	0.616						
-0.649	0.306	-0.052	-0.175	-0.846	-0.214	0.595						
-1.101	0.141	-0.040	-0.341	-1.372	0.042	0.738						
-0.799	0.225	-0.045	-0.233	-1.016	-0.100	0.624						
-3.973	-0.753	0.009	-1.380	-4.743	1.502	1.760						
-28.788	22.352	-3.399	-6.760	-38.970	-18.838	32.690						
-0.600	0.107	-0.622	0.578	-0.047	0.076	0.300						
-0.102	0.013	-0.399	0.209	-0.082	-0.287	0.359						
-0.684	0.121	-0.731	0.667	-0.059	0.065	0.283						
-0.331	0.056	-0.513	0.383	-0.069	-0.131	0.300						
-4.171	0.768	-2.818	3.449	0.051	2.077	0.890						
7.249	-1.666	-16.444	2.070	-5.532	-25.469	15.398						
-0.203	0.404	-0.118	-0.027	-0.256	-0.039	0.714						
-0.285	0.402	0.292	-0.169	-0.169	0.096	0.522						
-0.250	0.485	-0.112	-0.043	-0.301	-0.037	0.843						
-0.250	0.485	-0.112	-0.043	-0.301	-0.037	0.843						
-0.257	0.416	0.115	-0.111	-0.214	0.037	0.626						
-0.122	1.092	-2.432	0.654	-1.132	-0.802	2.903						
-18.085	22.431	26.988	-13.161	-7.181	8.896	24.150						
0.765	0.679	0.428	-0.489									
0.320	0.273	0.321	-0.206									
0.886	0.785	0.506	-0.567									
0.532	0.466	0.380	-0.341									
4.340	3.905	1.696	-2.770									
6.309	4.722	15.183	-4.147									

TABELLA 10

Parametri e matrici di regressione ottenuti mediante RCD

$$L_2^*(I - A)^{-1}P^* = \frac{1}{n} X_1^* S_1^{*-1} L_2^*(I - A)^{-1} = \frac{1}{n} XG^{**} S^{*-1} L_2^* = \frac{1}{n} YI^{**} S^{*-1} - \frac{1}{n} YG^{**} S^{*-1} = 0$$

$$\begin{bmatrix} 3.017 & 0 & 1.773 \\ 2.684 & 1.969 & -0.03 \\ 3.136 & -1.261 & -0.753 \\ 2.702 & -0.724 & -1.177 \end{bmatrix}$$

$$\begin{bmatrix} - & 1.973 \\ & -11.049 \end{bmatrix}$$

[0]

$$L = \frac{1}{n} R^* S^{*-1} R$$

$$\begin{bmatrix} 3.017 & 0 & 1.773 \\ 2.684 & 1.969 & -0.03 \\ 3.136 & -1.261 & -0.753 \\ 2.702 & -0.724 & -1.177 \\ -1.973 & 0 & 0 \\ -11.049 & 0 & 0 \end{bmatrix}$$

5. CONCLUSIONE

Le VL definite nel *soft modelling* di Wold hanno proprietà analoghe alle VL del LM e le loro *proxies* calcolate mediante il PLS permettono di ovviare ai problemi di identificabilità e indeterminazione propri del LM. Tuttavia tali *proxies*, oltre ad essere combinazioni lineari delle variabili osservate invece che "fattori causali" latenti, non possiedono molte altre proprietà delle VL del LM. Anche attraverso l'RCD si ottengono, in luogo che "fattori causali" latenti, componenti che sono combinazioni lineari delle variabili osservate: tuttavia tali componenti, oltre ad ovviare ai problemi di identificabilità e indeterminazione del LM, hanno per il resto proprietà del tutto analoghe alle VL del LM.

Istituto di Scienze statistiche e matematiche

"Marcello Boldrini"

Università degli Studi di Milano

GIORGIO VITTADINI

RIFERIMENTI BIBLIOGRAFICI

- H. APEL, H. WOLD (1982), *Soft modelling with latent variables in two or more dimensions: PLS estimation and testing for predictive relevance*, in K.G. Jöreskog & H. Wold (eds.), "Systems under indirect observation: Causality, structure, prediction", North Holland, Amsterdam, 2, pp. 209-247.
- K. HAAGEN (1991), *On necessary and sufficient conditions for vanishing the arbitrariness in a common factor model*, "The Statistician", 40, pp. 401-408.
- K. HAAGEN, G. VITTADINI (1991), *Regression component decomposition in structural analysis*, "Communications in Statistics", 20, pp. 1153-1161.

- B.S. HUI, H. WOLD, *Consistency and consistency at large of partial least squares estimates*, in K.G. Jöreskog & H. Wold (eds.), "Systems under indirect observation: Causality, structure, prediction", North Holland, Amsterdam, 2, pp. 119-130.
- K.G. JÖRESKOG (1970), *A general method for analysis of covariance structures*, "Biometrika", 57, pp. 239-251.
- K.G. JÖRESKOG (1971), *Simultaneous factor analysis in several populations*, "Psychometrika", 36, pp. 409-426.
- K.G. JÖRESKOG (1973a), *A general method for estimating a linear structural equation system*, in A.S. Goldberger & O.D. Duncan (eds.), "Structural equation models in the social sciences", Seminar Press, New York, pp. 85-112.
- K.G. JÖRESKOG (1973b), *Analysis of covariance structures*, in P.R. Krishnaiah (ed.), "Multivariate analysis", Academic press, New York, 3, pp. 263-285.
- K.G. JÖRESKOG (1974), *Analyzing psychological data by structural analysis of covariance matrices*, in R.C. Atkinson, D.H. Krantz, R.D. Luce, & P. Suppes (eds.), *Contemporary developments in mathematical psychology*, Freeman, San Francisco, 2, pp. 1-56.
- K.G. JÖRESKOG (1977), *Structural equation models in the social sciences: Specification, estimation and testing*, in P.R. Krishnaiah (ed.), "Applications of statistics", North Holland, Amsterdam, pp. 265-287.
- K.G. JÖRESKOG (1978), *Structural analysis of covariance and correlation matrices*, "Psychometrika", 43, pp. 443-477.
- K.G. JÖRESKOG (1979a), *Basic ideas of factor analysis and component analysis*, in K.G. Jöreskog & D. Sorbom (eds.), "Advances in factor analysis and structural equation models", Abt Books, Cambridge, Massachusetts, pp. 5-20.
- K.G. JÖRESKOG (1979b), *Statistical estimation of structural models in longitudinal developmental investigations*, in J.R. Nesselroade & P.B. Baltes (eds.), "Longitudinal research in the study of behavior and development", Academic Press, New York, pp. 303-351.
- K.G. JÖRESKOG (1981), *Basic issues in the application of LISREL*, "Data, communications computer data analysis", 1, pp. 1-6.
- K.G. JÖRESKOG (1982), *The LISREL approach to causal model-building in the social sciences*, in K.G. Jöreskog & H. Wold (eds.), "Systems under indirect observation: Causality, structure, prediction", North Holland, Amsterdam, 1, pp. 81-99.
- J.B. LOHMÖLLER (1989), *Latent variable path modelling with partial least squares*, Physica Verlag, Heidelberg.
- R. NOONAN, H. WOLD (1982), *PLS path modelling with indirectly observed variables: A comparison of alternative estimates for the latent variables*, in K.G. Jöreskog & H. Wold (eds.), *Systems under indirect observation: Causality, structure, prediction*, North Holland, Amsterdam, 2, pp. 75-94.
- P.H. SCHÖNEMANN, K. HAAGEN (1987), *On the use of factor scores for prediction*, "Biometrical Journal", 29, pp. 835-847.
- P.H. SCHÖNEMANN, J.H. STEIGER (1976), *Regression component analysis*, "British Journal of Mathematical and Statistical Psychology", 29, pp. 175-189.
- G. VITTADINI (1988), *On the validity of the indeterminate latent variables in the LISREL model*, "Communications in Statistics", 17, 3, pp. 861-874.
- G. VITTADINI (1989), *Indeterminacy problem in the LISREL model*, "Multivariate Behavioral Research", 24, 4, pp. 397-414.
- H. WOLD (1980), *Model construction and evaluation when theoretical knowledge is scarce: Theory and application of Partial Least Squares*, in J. Kmenta & J.B. Ramsey (eds.), "Model evaluation in econometrics", Academic, New York, pp. 47-74.