The PLS (Partial Least Squares) Approach to Multidimensional Contingency Tables

Wold, Herman; Bertholet, Jean-Luc

HERMAN WOLD (*) and JEAN-LUC BERTHOLET (**)

The PLS (Partial Least Squares) Approach to Multidimensional Contingency Tables

0. INTRODUCTION.

Background reference is made to the PLS approach to path models with latent variables, briefly called "soft modeling"; Wold (1975, 1977, 1979, 1980, 1982). The present paper shows that the basic design for PLS estimation of soft models allows straightforward adaptation to the analysis of the dichotomous items of multidimensional contingency tables. The ensuing model defines one or more latent variables for each margin of the contingency table; the observed items are interpreted as indicators of the corresponding latent variables; each latent variable is estimated explicitly as a weighted aggregate of its indicators; the model has "outer relations" between each latent variable and its indicators, and "inner relations" between the latent variables; the inner and outer relations are causal-predictive; for the indicators of a latent variable that is explained by an inner relation, substitutive elimination of the latent variable from the outer relations gives causal-predictive relations for the indicators in terms of the explanatory latent variables of the inner relation; the predictive relevance of any causal-predictive relation of the model can be tested by Stone-Geisser's test (1974), giving R^2 evaluated without loss of degrees of freedom. Several generalizations of the basic PLS estimation algo-

(*) Statistics Dept., University of Uppsala, and Econometrics Dept., Univ. of Geneva.

(**) Econometrics Dept., University of Geneva, and Sociological research service, Public education administration, Geneva.

rithm carry over to multidimensional contingency tables, including feedbacks in the inner relations; hierarchic structure of the latent variables; and latent variables in two or more dimensions.

Our paper has four sections:

1. PLS soft modeling: an overview of ends and means.
2. Contingency tables in two dimensions.
3. Contingency tables in three dimensions.
4. Further developments.

Lawrence Kohlberg's classical theory on moral evolution and his rich data bank have been restructured by Kurt Bergling (1981) for analysis by statistical methods of the ML (Maximum Likelihood) family. Kohlberg's data having the form of multidimensional contingency tables, it was our study of Bergling (1981) that led to the idea of using PLS instead of ML, and thereby to the present paper. To carry over PLS from scalar variables to the dichotomous items of multidimensional contingency tables is immediate matter, since PLS is distribution-free. What is new in the present paper, relative to PLS as applied to scalar variables, is the multiplicative combination of indicator items to form complex indicators of second or higher order. The reach and limitation of PLS in the analysis of complex indicators requires further comparative study of PLS versus other methods for the analysis of multidimensional contingency tables. We have every reason to expect that the following attractive features of PLS carry over from the analysis of scalar observables to the analysis of multidimensional contingency tables:

* Once the conceptual-theoretical design of the model is specified by an "arrow scheme", it is immediate matter to write down the formal model and the PLS estimation algorithm.
* The PLS algorithm is an iterative sequence of OLS (Ordinary Least Squares) regressions, and is therefore easy and speedy on the computer: "instant estimation".
* Hence the PLS approach can cope with quite large and complex models, and with massive data banks. For purposes of substantive analysis it is often necessary to work with complex models and large data banks.
* The PLS approach provides predictive inference, and the relevance of the predictions can be tested by Stone-Geisser's method.

In due course we plan to take up applied work with the PLS approach to multidimensional contingency tables, with emphasis on substantive analysis. It would be false modesty if we did not express the hope that the PLS approach will prove useful in the analysis of the classical theories and data banks of Jean Piaget and Lawrence Kohlberg.

1. PLS SOFT MODELING: AN OVERVIEW OF ENDS AND MEANS.

Soft modeling is primarily intended for multidisciplinary and other applications where the problems explored are complex and theoretical knowledge is scarce. In this section we shall briefly review the basic design of soft modeling with scalar variables. With reference to Figure 1 for illustration it will suffice for our purpose to consider a soft model with two blocks of manifest variables.

1.1 *Formal specification of the model.*

1.1.1 *Variables.* The model has two blocks of manifest variables (MVs), observed over N cases,

$$x_{hn}\,,\ y_{kn}\,,\quad h = 1\,,H\,;\quad k = 1\,,K\,;\quad n = 1\,,N \tag{1}$$

and two latent variables (LVs),

$$\xi_n\,,\quad \eta_n\,,\quad n = 1\,,N \tag{2}$$

In what follows the ranges of the subscripts will usually not be spelled out.

1.1.2 *Outer relations.* In each block the indicators are assumed to be linear in their LV:

$$x_{hn} = \pi_{1ho} + \pi_{1h}\,\xi_n + \varepsilon_{1hn}\,;\ y_{kn} = \pi_{2ko} + \pi_{2k}\,\eta_n + \varepsilon_{2kn} \tag{3}$$

As in factor analysis the multiplicative coefficients π_{1h}, π_{2k} are called the *loadings* of the indicators. Both π_{1h} and ξ_n being unknown, some standardization of scales in necessary to avoid ambiguity in the model, and similarly for π_{2k} and η_n. To achieve SSU (Standardization for Scale Unambiguity) in soft modeling, all LVs are standardized so as to have unit variance. Hence in the present case:

$$\text{var}\,\xi = E\,(\xi^2) - [E\,(\xi)]^2 = 1\,;\quad \text{var}\,\eta = E\,(\eta^2) - [E\,(\eta)]^2 = 1 \tag{4}$$

1.1.3 *Inner relations.* The present model has one inner relation:

$$\eta_n = \beta_0 + \beta_1 \xi_n + \upsilon_n \tag{5}$$

1.1.4 *Substitutive prediction* gives y_{kn} linearly in terms of ξ_n :

$$y_{kn} = \alpha_{ko} + \pi_{2k} \beta_1 \xi_n + \nu_{kn} \tag{6}$$

with location parameters (7a) and residuals (7b),

$$\alpha_{ko} = \pi_{2ko} + \pi_{2k} \beta_0 \; ; \quad \nu_{kn} = \varepsilon_{2kn} + \pi_{2k} \upsilon_n \tag{7}$$

1.1.5 *Comments.*

(i) The MVs are grouped in blocks, which are the structural units of the model.

(ii) In each block the MVs are assumed to be indicators of an LV.

(iii) The grouping of the observables into blocks is instrumental in reducing the complexity of the model, and so is the introduction of LVs.

(iv) The outer relations between each LV and its indicators, and the inner relations between the LVs constitute the formal definition of the model.

(v) The arrow scheme illustrates the inner and outer relations by directed arrows. The arrows indicate *channels of information* in the model.

(vi) All information between the blocks is assumed to be conveyed by the LVs. Accordingly, the residual of any outer relation is assumed to be uncorrelated with the LVs as well as with the residual of any outer relation in the other block.

(vii) The arrow scheme marks the residual of any inner or outer relation by an arrow head; the residuals are not illustrated otherwise in the arrow scheme of a soft model.

1.2 *PLS estimation.*

The PLS algorithm proceeds in three stages. In the two first stages each indicator is measured by the deviations from its mean, giving

$$\bar{x}_h = 0 \,, \quad \bar{y}_k = 0 \,, \quad h = 1 \,, H \,; \quad k = 1 \,, K \tag{8}$$

The third stage estimates the location parameters of the model.

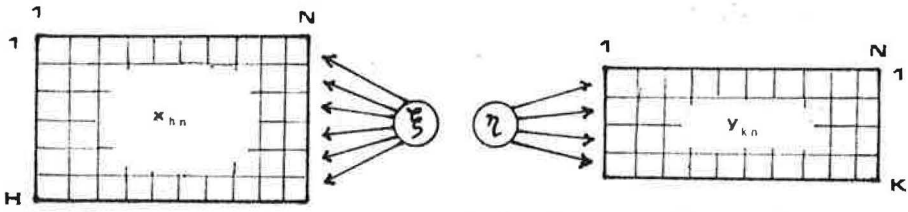


Fig. 1. – Arrow scheme for a soft model with two blocks of observables x_h, x_k and two latent variables ξ, η.

1.2.1. The *first stage* of the PLS algorithm performs an explicit estimation of the LVs, each LV being estimated as a weighted aggregate of its indicators:

$$X_n = \text{est}\ \xi_n = f_1\ \Sigma_h\ (w_{1h}\ x_{hn})\ ;\quad Y_n = \text{est}\ \eta_n = f_2\ \Sigma_k\ (w_{2k}\ y_{kn}) \tag{9}$$

where (8) gives

$$\bar{X} = \frac{f_1}{N}\ \Sigma_h\ (w_{1h}\ \bar{x}_h) = 0\ ;\quad \bar{Y} = \frac{f_2}{N}\ \Sigma_k\ (w_{2k}\ \bar{y}_k) = 0 \tag{10}$$

and f_1, f_2 are scalars determined so as to give X_n and Y_n unit variance, in accordance with (10):

$$\frac{1}{N}\ \Sigma_n\ (X_n^2) = 1\ ,\quad \frac{1}{N}\ \Sigma_n\ (Y_n^2) = 1 \tag{11}$$

The *weights* w_{1h}, w_{2k} are determined by the *weight relations*:

$$x_{hn} = w_{1h}\ Y_n + d_{1hn}\ ;\quad y_{kn} = w_{2k}\ X_n + d_{2kn} \tag{12}$$

In the first stage the PLS is iterative, alternating between (9) and (12), using the starting values:

$$w_{11} = w_{21} = 1 \text{ and } w_{1h} = w_{2k} = 0 \text{ for } h = 2\ ,\ H\ ;\ k = 2\ ,\ K \tag{13}$$

For two-block, two-LV models the iterative procedure converges almost certainly (unit probability); cf. Lyttkens, Areskoug and Wold (1975).

1.2.2 *The second stage* of the PLS algorithm estimates the inner and outer relations by corresponding OLS regressions, using the LVs estimated in the first stage.

Outer relations:

$$x_{hn} = p_{1h}\ X_n + e_{1n}\ ;\quad y_{kn} = p_{2k}\ Y_n + e_{2n} \tag{14}$$

Inner relation:

$$Y_n = b_1 X_n + u_n \tag{15}$$

Substitutive prediction of the indicators y_{kn} in terms of X_n :

$$y_{kn} = p_{2k} b_1 X_n + v_{kn} ; \quad v_{kn} = e_{2n} + p_{2k} u_n \tag{16}$$

Substitutive prediction of the indicators y_{kn} in terms of the indicators x_{hn} :

$$y_{kn} = p_{2k} b_1 f_1 \Sigma_h (w_{1h} x_{hn}) + v_{kn} \tag{17}$$

1.2.3 *The third stage* of the PLS algorithm, cancelling the standardization (8) to zero means, estimates the location parameters of the LVs and the relations. This is immediate matter, as always in OLS regressions. Thus for (9b), (14b), (15) and (16a) the location parameters are

$$\bar{Y} = f_2 \Sigma_k (w_{2k} \bar{y}_k) ; \quad p_{2ko} = \bar{y}_k - p_{2k} \bar{Y} ; \tag{18}$$

$$b_0 = \bar{Y} - b_1 \bar{X} ; \quad a_{ko} = p_{2ko} + p_{2k} b_0 . \tag{19}$$

1.3 *Testing for predictive relevance.*

Using Stone-Geisser's test, Wold (1982) explains the procedure in detail for the outer and inner relations. Let us briefly restate the procedure for the inner relation (15).

The LVs are regarded as directly observed by their estimates X_n, Y_n ; the test proceeds in G rounds, having chosen for G an integer in the range $10 \leq G \leq 15$; in the g^{th} round ($g = 1, G$) the LVs

$$Y_g , Y_{g+G} , Y_{g+2G} , \ldots \tag{20}$$

are removed from the data; we predict the removed LVs by

$$\text{pred } Y_{g+rG} = b_{2o} + b_{21} X_{g+rG} ; \quad r = 0, 1, 2, \ldots . \tag{21}$$

and form the sum of squares of the prediction errors,

$$S S^{(g)} = \Sigma_r (Y_{g+rG} - \text{pred } Y_{g+rG})^2 ; \tag{22}$$

comparing with the trivial predictions

$$\text{triv pred } Y_{g+rG} = \frac{1}{N-1} \Sigma_{n \neq g+rG} (Y_n) \tag{23}$$

and the corresponding square sum of errors,

$$\text{triv } SS^{(g)} = \Sigma_r \, (Y_{g+rG} - \text{triv pred } Y_{g+rG})^2 \tag{24}$$

we obtain Stone-Geisser's test criterion:

$$Q^2 = 1 - \Sigma_g \, (SS^{(g)}) \, / \, \Sigma_g \, (\text{triv } SS^{(g)}) \, . \tag{25}$$

If the inner relation has predictive relevance, $Q^2 > 0$. Lack of predictive relevance is revealed by $Q^2 < 0$. In bordering cases, $Q^2 \sim 0$, the decision is uncertain.

2. Contingency tables in two dimensions.

We shall now set forth the PLS approach to contingency tables, beginning with the special case of tables with two margins. Later in this section we shall give numerical examples, and adduce some comments.

2.1 *Notation and statistical hypotheses.*

2.1.1 *Manifest and latent variables.* With reference to Figure 2 for illustration, let Δ denote a contingency table with H rows and K columns. Δ_{hk} denotes the absolute frequency of the hth characteristic of the first qualitative variable, and the kth of the second. Let N denote the total number of observations (cases):

$$N = \sum_{h=1}^{H} \sum_{k=1}^{K} \Delta_{hk} \, . \tag{26}$$

Each of the N observations takes the form of two column vectors:

$$\underset{H\times 1}{x_1} \, , \quad \underset{K\times 1}{x_2} \tag{27}$$

In each vector all entries are zero except for a unit entry at the row that indicates the characteristic of the observation. For example, if an observation shows the third characteristic of x_1 and the second of x_2, the vectors (27) when transformed to row vectors will read:

$$x_1' = \| x_{1h} \| = (0\,0\,1\,0\,0 \ldots 0) \, , \; h = 1 \, , H \, ; \; x_2' = \| x_{2k} \| = (0\,1\,0\,0 \ldots 0) \, , \; k = 1 \, , K \tag{28}$$

where the prime (′) denotes transposition.

2.1.2 *The arrow scheme.* Again with reference to Figure 2, the arrow scheme constitutes the conceptual-theoretical model. In the present case the model has two LVs:

$$\xi_{1n}\,, \quad \xi_{2n} \quad n = 1\,, N \tag{29}$$

The two margins in Figure 2 correspond to the two blocks of indicators in Figure 1. Hence the PLS analysis of the arrow scheme in Figure 1 carries over directly to the arrow scheme in Figure 2.

2.1.3 *Outer relations.* In the basic design of PLS soft modeling the outer relations are assumed to be linear, in the present case:

$$x_{jkn} = \pi_{jko} + \pi_{jk}\,\xi_{jn} + \varepsilon_{jkn}\,, \quad j = 1\,, 2\,; \quad k = 1\,, K_j\,; \quad n = 1\,, N \tag{30}$$

where $K_1 = H\,;\ K_2 = K\,.$

The outer relations are subject to predictor specification:

$$E\,(x_{jk} \mid \xi_j) = \pi_{jko} + \pi_{jk}\,\xi_j \quad j = 1\,, 2 \tag{31}$$

The predictor specification implies the corollaries (32 a-b); in words:

* Each residual has conditional expectation zero;

** Each residual is uncorrelated with the LV;

$$E\,(\varepsilon_{jk} \mid \xi_j) = 0\,, \quad r\,(\varepsilon_{jk}\,, \xi_j) = 0\,, \quad j = 1\,, 2\,; \quad k = 1\,, K_j \tag{32}$$

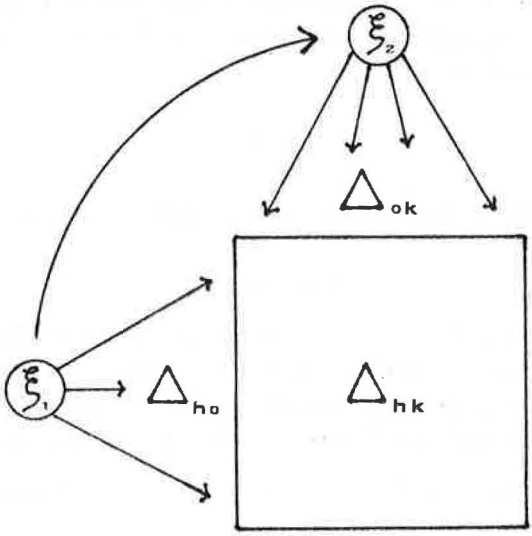


Fig. 2. – Arrow scheme for PLS modeling of a contingency table with two margins.

To achieve SSU in the outer relations (30), cf. (4), the scales of the LVs are standardized to give the LVs unit variance:

$$\text{var } \xi_j = 1 , \quad j = 1 , 2 \tag{33}$$

To repeat from 1.1.5 (vi), it is a fundamental principle in soft modeling that the information between the blocks, and the ensuing causal-predictive inference, is conveyed by the LVs. Accordingly, it is assumed that the LVs in general are intercorrelated, in the present case say

$$r (\xi_1 , \xi_2) = \rho_{12}$$

whereas the residuals of any block are assumed to be uncorrelated with the residuals and the LV of the other block; that is,

$$r (\varepsilon_{1h} , \varepsilon_{2k}) = 0 ; \quad r (\varepsilon_{1h} , \xi_2) = 0 ; \quad r (\varepsilon_{2k} , \xi_1) = 0 \tag{34}$$

Here, the residuals are mutually correlated within the blocks because of the restriction $\sum_{k=1}^{K_j} x_{jkn} = 1$, for all j and n, cf. eq. (28). This gives

$$r (\varepsilon_{1h} , \varepsilon_{1h'}) \neq 0 , h , h' = 1 , H ; \quad r (\varepsilon_{2k} , \varepsilon_{2k'}) \neq 0 , k , k' = 1 , K \tag{35}$$

Eq. (35) does not interfere with PLS estimation algorithm but is of relevance of the dimensionality of the LVs; cf. Apel and Wold (1982).

2.1.4 *Inner relation.* The present model has one inner relation, and this is assumed to be linear,

$$\xi_{2n} = \beta_0 + \beta_1 \xi_{1n} + \upsilon_n \tag{36}$$

and is subject to predictor specification,

$$E (\xi_2 \mid \xi_1) = \beta_0 + \beta_1 \xi_1 \tag{37}$$

2.1.5 *Substitutive prediction.* Eliminating ξ_2 from (30b) by means of (36) we obtain x_{2k} linearly in terms of ξ_1 :

$$x_{2kn} = \alpha_{2ko} + \pi_{2k} \beta_1 \xi_{1n} + \nu_{2kn} \tag{38}$$

where

$$\alpha_{2ko} = \pi_{2ko} + \pi_{2k} \beta_0 ; \quad \nu_{2kn} = \varepsilon_{2k} + \pi_{2k} \upsilon_n \tag{39}$$

2.2 PLS *estimation.*

Introducing the matrices Z_1, Z_2 by

$$\underset{N\times H}{Z_1} = \begin{pmatrix} x'_{11} \\ \vdots \\ x'_{1N} \end{pmatrix}; \quad \underset{N\times K}{Z_2} = \begin{pmatrix} x'_{21} \\ \vdots \\ x'_{2N} \end{pmatrix} \tag{40}$$

we obtain

$$\Delta = Z'_1 Z_2 \tag{41}$$

The two first stages of PLS algorithm work with indicators, say x^*_{1h}, x^*_{2k}, that are standardized to zero mean, giving

$$\bar{x}^*_{1h} = 0\,, \qquad \bar{x}^*_{2k} = 0\,, \qquad h = 1\,, H\,; \quad k = 1\,, K \tag{42}$$

The standardized indicators take the form

$$\underset{N\times H}{x^*_1} = H Z_1\,, \qquad \underset{N\times K}{x^*_2} = H Z_2 \tag{43}$$

where H is the idempotent matrix defined by

$$H = [I_N - N^{-1} L_N L'_N]\,, \quad (H = H^2) \tag{44}$$

writing I_N for the unity matrix, and L_N for a $(N \times 1)$ column vector of units.

2.2.1 *First stage of the PLS estimation procedure.* The LVs ξ_1, ξ_2 are estimated by weighted aggregates of their indicators, say X_{1n}, X_{2n}. Denoting the *weights* by

$$w'_1 = (w_{11}\,, \ldots\,, w_{1H})\,, \quad w'_2 = (w_{21}\,, \ldots\,, w_{2K}) \tag{45}$$

the estimated LVs take the form

$$X_1 = \text{est}\ \xi_1 = f_1 H Z_1 w_1\,; \quad X_2 = \text{est}\ \xi_2 = f_2 H Z_2 w_2 \tag{46}$$

where f_1, f_2 are scalars that standardize X_1 and X_2 to unit variance,

$$f_j = (w'_j Z'_j H Z_j w_j)^{-1/2}\,, \quad j = 1\,, 2 \tag{47}$$

The *weight relations* given by

$$x^*_{1h} = w_{1h} X_2 + d_{1h}\,; \qquad x^*_{2k} = w_{2k} X_1 + d_{2k} \tag{48}$$

serve to determine the weights. We can write (48) in a more compact form:

$$H Z_1 = X_2 w_1' + D_1 ; \quad H Z_2 = X_1 w_2' + D_2 \tag{49}$$

where

$$\underset{N \times H}{D_1} = (d_{11}, \ldots, d_{1H}) ; \quad \underset{N \times K}{D_2} = (d_{21}, \ldots, d_{2K}) \tag{50}$$

* The first stage is iterative, say with steps $s = 1, 2, \ldots$, and alternates in each step between (46) and (49). In the start, $s = 1$, there is an almost free choice of the weights, say $w_1 = w_2' = (1, 0, 0, \ldots, 0)$. Having obtained the weights for step s, (46) gives X_1 and X_2 in step s; then the simple OLS regressions (49) give the weights for step $s + 1$. The procedure continues until the weights stop changing between two steps according to some standard criterion.

2.2.2 *Second stage of the PLS estimation procedure.* Using the LVs estimated in the first stage, the outer and inner relations are estimated by corresponding OLS regressions. Theoretical and estimated parameters and residuals are denoted by corresponding Greek and Roman letters.

Outer relations, cf. eq. (30):

$$x_{1h}^* = p_{1h} X_1 + e_{1h} ; \quad x_{2k}^* = p_{2k} X_2 + e_{2k} \tag{51}$$

The loadings p_{1h}, p_{2k} measure the relations between an LV and its indicators, and are thereby analogous to the loadings of classical factor analysis. The PLS loadings are simple OLS regression coefficients:

$$p_{1h} = X_1' x_{1h}^* = f_1 w_1' Z_1' H x_{1h} = f_1 \Delta_{h\cdot} [w_{1h} - S_1] \tag{52}$$

where

$$S_1 = N^{-1} \Sigma_{h=1}^{H} (\Delta_{h\cdot} w_{1h}) , \tag{53}$$

and similarly for p_{2h}. — We see that the various entries of the contingency table are not needed: the margins are sufficient.

Inner relation, cf. eq. (36):

$$X_2 = b_1 X_1 + u \tag{54}$$

The inner parameter b_1 is equal to the correlation of X_1 and X_2:

$$b_1 = r (X_1, X_2)$$

$$N b_1 = X_1' X_2 = w_1' Z_1 H Z_2 w_2 = w_1' [\Delta - \Delta_1 \Delta_2' N^{-1}] w_2 \tag{55}$$

where

$$\Delta_1 = \Delta\, L_H\,; \quad \Delta_2 = \Delta'\, L_K \tag{56}$$

Substitutive prediction. The causal-predictive relations (6) - (7) carry over to contingency tables; to spell out the counterpart to (16):

$$x^*_{2kn} = p_{2k}\, b_1\, X_{1n} + e_{2kn} + p_{2k}\, u_n \tag{57}$$

2.2.3 The *third stage* of the PLS algorithm estimates the location parameters.

Formulas (18) - (19) carry over; for example, the location parameters of X_2 and the outer relation (51 *b*) are given by

$$\overline{X}_2 = f_2\, w_2'\, \overline{x}_2 = f_2\, \Sigma_k\, (w_{2k}\, \overline{x}_{2k})\,; \quad p_{2ko} = \overline{x}_{2k} - p_{2k}\, \overline{X}_2 \tag{58}$$

2.3 *Numerical illustrations.*

We shall briefly present and then discuss three simple examples of twodimensional contingency tables. In all three examples the arrow scheme is of the same type as in Figure 2. A programme witten by Jan-Bernd Lohmöller was used for all the following numerical examples.

2.3.1 *Example* 1: See Table 1. This artificial example gives a one-to-one correspondence between the rows and columns of a diagonal contingency table.

TABLE 1

PLS ANALYSIS OF DIAGONAL CONTINGENCY TABLES WITH TWO MARGINS

DATA

2	0	0	2
0	4	0	4
0	0	6	6
2	4	6	12

PLS estimates

Block	Variable	Weight	Loading
row	row 1	— .1553	— .0630
	row 2	— 1.0243	— .4156
	row 3	1.1796	.4786
column	column 1	— .1552	— .0629
	column 2	— 1.0244	— .4156
	column 3	1.1796	.4786

Since the table is diagonal, the PLS loadings are proportional to the PLS weights; $b_1 = r(X_1, X_2) = 1$.; and $R^2 = 1$ for the inner relation.

2.3.2 *Example* 2: A 3 × 2 contingency table; see Table 2.

In this contingency table the first row is independent of the other ones in the sense that it is proportional to the row margin. As a consequence, the corresponding weight is zero, $w_{11} = 0$.

TABLE 2

PLS ANALYSIS OF A 3 × 2 CONTINGENCY TABLE

DATA

2	4	6
1	3	4
5	9	14
8	16	24

PLS estimates

Block	Variable	Weight	Loading
row	row 1	.0000	— .1372
	row 2	— 1.3172	— .3110
	row 3	1.3172	.4482
column	column 1	— 1.0607	— .4714
	column 2	1.0607	.4714

The correlation between the two LVs is very low,

$$r(X_1, X_2) = b_1 = -.078$$

which shows that the two blocks of indicators are almost independent, and that the R^2 of the inner relation is very small, $R^2 = .006$.

2.3.3 *Example* 3. A 2 × 2 contingency table based on real-world data; see Table 3.

TABLE 3

PLS ANALYSIS OF A 2 × 2 CONTINGENCY TABLE; DATA FROM STOUFFER ET AL. (1949)

DATA

	Preference N	Preference S	
present N	18	6	24
location S	22	33	55
	40	39	79

PLS estimates

Block	Variable (code)	Weight	Loading
present location	North	— 1.0872	— .4599
	South	1.0872	.4599
preference	North	— 1.0001	— .5000
	South	1.0001	.5000

$b_1 = r(X_1, X_2) = .322$; $R^2 = .1037$

This example draws from extensive data collected by Stouffer et al. (1949), and analyzed by many researchers. Our simple 2×2 table is a cross classification of the present location of the soldiers (North, South) and their preference as to camp location (North, South); the unit is 1,000 soldiers.

In the vertical margin we see that South dominates the present location, whereas the horizontal margin shows that North and South balance in the preferred location. The degree of similarity between present and preferred location is measured by $r(X_1, X_2) = b_1 = .32$.

2.4 *Comments.*

(i) Model building with latent variables, and their explicit PLS estimation, are key features in soft modeling with scalar variables, and are novel features in the analysis of muldidimensional contingency tables.

(ii) The inner and outer relations (3) - (5) and the substitutive predictions (6) - (7) of a soft model with scalar variables carry over to the PLS approach to multidimensional contingency tables, where they constitute novel modes of inference.

(iii) The explicit PLS estimation of the latent variables of a soft model is deliberately approximate. The limiting PLS estimates of the latent variables are inconsistent (biased in the large-sample sense), and so are the ensuing PLS parameter estimates of inner and outer relations and of substitutive prediction.

(iv) The PLS estimates of latent variables and parameters are consistent at large; that is, with increasing numbers of indicators for each latent variable the estimates will under general conditions of regularity tend to be consistent.

(v) For the estimation of soft models with two latent variables the iterative PLS estimation of the latent variables will almost certainly converge (unit probability), and be invariant to the choice of starting values; cf. 2.2.1*.

(vi) Thanks to the explicit estimation of the latent variables, no identification problems arise in PLS soft modeling.

(vii) The zero correlation assumptions (35) imply that each latent variable has just one dimension. Whether or not the assumptions (35)

are fulfilled, the PLS estimation algorithm will give the first dimension of each latent variable.

(vii) In PLS soft modeling with scalar variables the investigator has the option to choose between two types of weight relations, called Mode A and Mode B. The weight relations (12) and (38) are Mode A; the PLS approach to multidimensional contingency tables Mode B is not applied here.

(ix) In the PLS approach to multidimensional contingency tables several features emerge that in general do not carry over to PLS soft modeling with scalar variables. Among those:

$$\Sigma_h w_{1h} = \Sigma_k w_{2k} = 0 \; ; \quad \Sigma_h p_{1h} = \Sigma_k p_{2k} = 0 \; ; \tag{59}$$

in words: the weights of any latent variable sum up to zero, and the same for the loadings. For illustrations, see Tables 1-3, and Tables 5-6 in Section 3.

(x) In the present paper there are several loose ends to tie up. For one thing, we have carried through only the first two stages of the PLS estimation procedure; that is, we have ignored all location parameters.

3. Contingency tables in three dimensions.

The PLS analysis of contingency tables with more than two dimensions is in line with what we have seen in Sn. 2, and is a straightforward adaptation of corresponding soft models with scalar variables. The more dimensions in the contingency table, the more numerous are the available designs of the arrow scheme.

3.1 *Data and arrow schemes.* Table 4 shows a three-dimensional contingency table that draws from Stouffer et al. (1949) to add a third dimension in the table analyzed in Example 3.

Figure 3 shows two arrow schemes, called Models A and B, that we shall use in the PLS analysis of Table 4. In Model A the indicators x_{1g}, x_{2h} in two margins define ξ_1 and ξ_2, two LVs that are assumed to influence the LV 3 that has its indicators x_{3k} in the third margin. In Model B we take account of the possible interaction between the indicators x_{1g} and x_{2h}, thereby defining a latent variable, denoted ξ_{12}, which joins ξ_1 and ξ_2 in influencing ξ_3. As applied to the data in Table 4, preference for camp

TABLE 4

Origin		Present location		Origin × present location		Preference		
North	South	North	South	code	freq.	North	South	
*		*		NN	15	13	2	
*			*	NS	26	18	8	
	*	*		SN	9	5	4	
	*		*	SS	31	5	26	
41	40	24	57			41	40	81

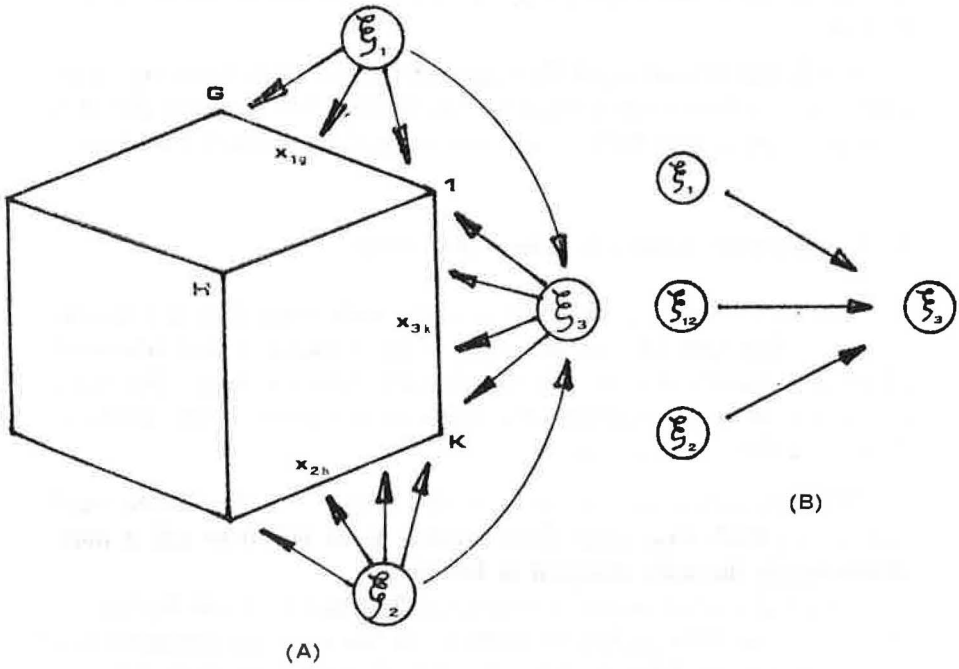


Fig. 3 A-B. – Two models for PLS analysis of a three-dimensional contingency table. A: The indicators of each margin define a latent variable; the two first latent variables, ξ_1 and ξ_2, influence the third, ξ_3. — B: The interaction between indicators x_{1g} and x_{2h} define a latent variable, denoted ξ_{12}, which joins ξ_1 and ξ_2 in influencing ξ_3.

location is our predicted LV ξ_3, while origin and present location are our explanatory LVs ξ_1 and ξ_2.

3.2 *PLS Models A and B applied to Table* 4.

The ensuing PLS estimates are shown in Tables 5 and 6.

All through, the LVs are estimated by weighted aggregates of type (46). As to the first stage of the PLS algorithm we spell out the weight relations for Model B:

$$\begin{aligned} H\,Z_1 &= X_3\,w_1' + D_1 \\ H\,Z_2 &= X_3\,w_2' + D_2 \\ H\,Z_3 &= (s_{3,1}\,X_1 + s_{3,2}\,X_2 + s_{3,12}\,X_{12})\,w_3' + D_3 \end{aligned} \tag{60}$$

where $s_{3,c}$ is $+1$ or -1 according as $r(X_3, X_c)$ is positive or negative.

All outer relations take the form (51). For Model B the inner relation reads:

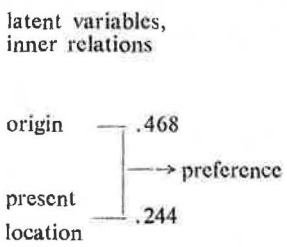
$$\xi_3 = \beta_{30} + \beta_{31}\,\xi_1 + \beta_{32}\,\xi_2 + \beta_{3,12}\,\xi_{12} + \upsilon_3 \tag{61}$$

The inner relation for Model A is obtained by omitting the term $\beta_{3,12}\,\xi_{12}$.

As always in PLS soft modeling the LVs are standardized to unit variance; hence the structural parameters of Models A and B can be readily compared. In both models the present location has less influence

TABLE 5

PLS ANALYSIS OF MODEL A



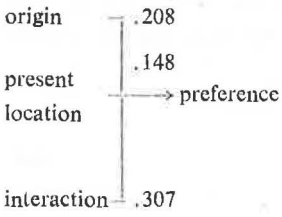
Block	Variable (code)	Weight	Loading
origin	North	1.0001	.5000
	South	— 1.0001	— .5000
present location	North	1.0950	.4566
	South	— 1.0950	— .4566
preference	North	1.0001	.5000
	South	— 1.0001	— .5000

$R^2 = .314$; $r(X_1, X_2) = .154$; $r(X_1, X_3) = .506$; $r(X_2, X_3) = .306$

TABLE 6

PLS ANALYSIS OF MODEL B

latent variables, inner relations



Block	Variable (code)	Weight	Loading
origin	North	1.0001	.5000
	South	— 1.0001	— .5000
present location	North	1.0950	.4566
	South	— 1.0950	— .4566
interaction	NN	.7326	.1730
	NS	.6556	.2751
	SN	.0602	.0291
	SS	— 1.4484	— .4772
preference	North	1.0001	.5000
	South	— 1.0001	— .5000

$R^2 = .324$

	Correlations of latent variables			
	Origin	Present location	Interaction	Preference
origin	1.000	.154	.896	.506
present location		1.000	.443	.316
interaction			1.000	.559
preference				1.000

on camp location than the soldier's origin. Model B shows that the interaction effect of origin and present location (.307) is even more important than the two separate effects (.208 and .148).

We clearly see in the data that present location has relatively small effect for soldiers whose region of origin is North, while it has substantial influence for soldiers from the South. In the same vein, most of the soldiers with origin in the North still prefer North when they are in the South, but soldiers originating from the South often prefer the North when they are located there. That is, interaction of origin and present location must not be neglected in the analysis.

3.3 *Validation of the model.*

3.3.1 The Stone-Geisser test was mentioned in the Introduction as a general test for predictive relevance in PLS soft modeling. We shall now present a specific validation technique for PLS analysis of contingency tables which uses the Substitutive prediction (57) for direct confrontation of model v.s. data. The classical validation methods usually perform a comparison of two tables: one is the original data table (in the present case Δ), while the other is an approximation of Δ (denoted $\hat{\Delta}$), which is constructed from the model and its hypotheses. The differences between the two tables are evaluated by some overall criterion, for example the Chi-square test.

3.3.2 To repeat, the PLS approach is prediction-oriented. By our Model A we try to predict prefecence for camp location by means of origin and present location. As specified in Table 7 these two explanatory variables determine four rows where the model must allocate and separate soldiers with preference for North vs. South.

Starting from raw data on (Z_1, Z_2, Z_3), the table Δ is computed by summing rows of Z_3 that correspond to the observations inside the cell; the allocation of an observation to a cell depends on information contained in Z_1 and Z_2. The proxy table $\hat{\Delta}$ will be computed in the same way; therefore we must first approximate Z_3 by $\hat{Z}_3$. This can easily be done by means of the outer and inner relations.

TABLE 7

NOTATION FOR CELL FREQUENCIES IN PREDICTIVE VALIDATION OF MODEL A

Explanatory variables			Predicted variables			
Code		Frequencies	Observed frequencies		Estimated frequencies	
Origin	Present location	Origin × present loc.	Preference		Preference	
			North	South	North	South
N	N	$n_{11.}$	n_{111}	n_{112}	$\hat{n}_{111}$	$\hat{n}_{112}$
N	S	$n_{12.}$	n_{121}	n_{122}	$\hat{n}_{121}$	$\hat{n}_{122}$
S	N	$n_{21.}$	n_{211}	n_{212}	$\hat{n}_{211}$	$\hat{n}_{212}$
S	S	$n_{22.}$	n_{221}	n_{222}	$\hat{n}_{221}$	$\hat{n}_{222}$

The problem before us belongs under Substitutive prediction, namely to estimate Z_3 in terms of X_1 and X_2; cf. (38) and (57). The outer and inner relations for the third LV of Model A are:

$$H Z_3 = X_3 p_3' + e_3 \tag{62a}$$

$$X_3 = b_{31} X_1 + b_{32} X_2 + u_3 \tag{62b}$$

Using Z_3 instead of $H Z_3$ by taking account of the location parameter, we obtain

$$Z_3 = X_3 p_3' + L_N N^{-1} \Delta_3' + e_3 \tag{63}$$

where

$$\Delta_3 = Z_3' L_N \tag{64}$$

Now for X_3 in (63) we substitute its prediction from (62b); that is:

$$X_3 = b_{31} X_1 + b_{32} X_2 \tag{65}$$

which gives, denoting predicted Z_3 by $\hat{Z}_3$

$$\hat{Z}_3 = (b_{31} X_1 + b_{32} X_2) p_3' + L_N N^{-1} \Delta_3' . \tag{66}$$

Recalling that X_1 and X_2 are aggregates of Z_1 and Z_2, respectively, eq. (66) implies that $\hat{Z}_3$ is a function of Z_1 and Z_2.

Note that $\hat{Z}_3$ fulfils two important constraints of Z_3, namely:

$$Z_3 L_2 = L_N , \qquad \hat{Z}_3 L_2 = L_N \tag{67}$$

$$L_N' Z_3 = \Delta_3' , \qquad L_N' \hat{Z}_3 = \Delta_3' \tag{68}$$

Hence:

$$L_N' Z_3 L_2 = N , \qquad L_N' \hat{Z}_3 L_2 = N \tag{69}$$

Eqs. (67) - (69) are interesting in that when computing $\hat{\Delta}$ from $\hat{Z}_3$ we are sure that the summations over each row and column of $\hat{\Delta}$ will give the same sums as for Δ.

3.3.3 *Numerical example.* PLS estimation of Model A has given the following results:

$$X_3 = .468 X_1 + .244 X_2 + u_3 \tag{70}$$

$$H Z_3 = X_3 p_3' + e_3 , \text{ with } p_3' = (-.5000 , +.5000) \tag{71}$$

$$H \hat{Z}_3 = (.468 X_1 + .244 X_2) p_3' \tag{72}$$

The ensuing comparison of observed and estimated cell frequencies is set forth in Table 8.

TABLE 8

PREDICTIVE VALIDATION OF MODEL A: NUMERICAL RESULTS

Explanatory variables			Predicted variable			
Code		Frequencies	Observed frequencies		Estimated frequencies	
Origin	Present location	Origin × present loc.	Preference		Preference	
			North	South	North	South
N	N	15	13	2	13.8810	1.1190
N	S	26	18	8	17.1132	8.8868
S	N	9	5	4	4.1166	4.8834
S	S	31	5	26	5.8962	25.1038
			n_{gh1}	n_{gh2}	$\hat{n}_{gh1}$	$\hat{n}_{gh2}$
		Sums:	41	40	41.0060	39.9930

4. DISCUSSION.

4.1 With reference to the *correspondence analysis* of twodimensional contingency tables introduced by Benzécri (1973), it will be noted that predictive inference is a common denominator of correspondence analysis and the PLS approach to contingency tables.

Let Δ_{hk} an $H \times K$ contingency table with population data on origin $(h = 1, H)$ versus preferred location $(k = 1, K)$. Benzécri transforms the data by

$$\Delta^*_{hk} = (\Delta_{hk} - N^{-1} \Delta_{h.} \Delta_{.k}) / (\Delta_{h.} \Delta_{.k})^{-1/2} \tag{73}$$

and computes the first principal component of Δ^*_{hk}, giving

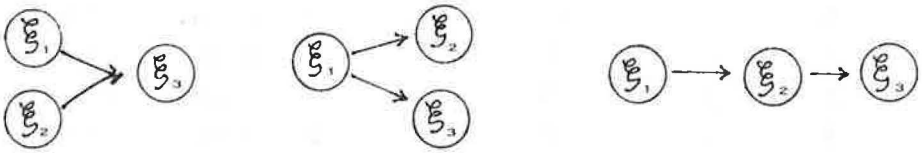
$$\Delta^*_{hk} = p_h X_k + e_{hk} \tag{74}$$

where p_h denotes the loadings, and X_k the component scores. Thus (74) predicts Δ^*_{hk} by $p_h X_k$, with prediction error e_{hk}.

The first principal component is a special case of the PLS algorithm; cf. Wold (1966, 1982). In the PLS approach the first principal component

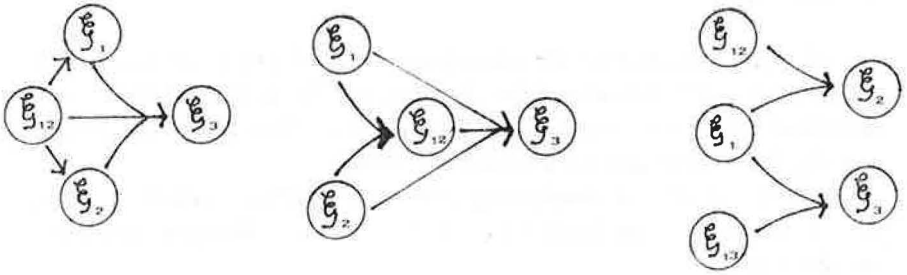
is conceived of as the estimate of a latent variable. From this point of view (74) is a prediction formula for the entries of a twodimensional contingency table, whereas Sn. 2 of the present paper gives prediction formulas in terms of two latent variables.

4.2 The basic design of the PLS algorithm provides estimation of path models with latent variables, and covers any design of the path of the (inner) relations between the latent variables. As applied to multidimensional contingency tables the investigator can choose between a variety of designs for the inner relations. The model in Figure 3A has the subsequent path (a) with just one inner relations, whereas the paths (b)



and (c) have two inner relations.

Combining the marginal indicators to form complex latent variables there is a rapid increase in the number of available designs, among those Figure 3B and many more, such as:



4.3 The basic design of PLS soft modeling has been generalized in several respects, which gives opportunities for corresponding developments of the models in Sn. 4.2, including: hierarchic structure of the latent variables, Wold (1982); latent variables in two or more dimensions, Apel and Wold (1982); and multiway observation of the indicators, Lohmöller (1981), Lohmöller and Wold (1980).

Clearly, the PLS approach to multidimensional contingency tables bristles with novel models to explore, to apply, and to compare with other approaches.

ACKNOWLEDGEMENTS

It is gratefully acknowledged that the present paper belongs under a grant from the Stiftung Volkswagenwerk in support of research on PLS soft modeling.

We are indebted to Jan-Bernd Lohmöller, M.A., Hochschule der Bundeswehr, Munich, and Professor Jack McArdle, University of Denver, for reading and commenting on a draft for our paper.

REFERENCES

APEL, H. and WOLD, H. (1982). *Higher dimensions for the latent variables in soft modelling, and testing for predictive relevance.* Chapter 10 in Wold (ed. 1982).

BENZÉCRI, J.-P. (1973). *L'analyse des données*, Vol. 2: *L'analyse des correspondances*, Paris, Dunod.

BERGLING, K. (1981). *Moral Development. The Validity of Kohlberg's Theory.* Acta Universitatis Stockholmiensis, Stockholm Studies in Educational Psychology 23.

GEISSER, S. (1974). *A predictive approach to the random effect model.* Biometrika, 61, 101-107.

GOODMAN, L.A. (1978). *Analyzing Qualitative/Categorical Data.* Log-Linear Models and Latent Structure Analysis. Cambridge Mass., Abt Books.

KOHLBERG, L. (1968). *Moral Development.* Volume 10, pages 483-494 in International Encyclopedia of the Social Sciences, ed. D.L. Sills. New York, MacMillan and Free Press.

LOHMÖLLER, J.-B. (1983). *Path models with latent variables and Partial Least Squares (PLS) estimation*, Doctoral Thesis, Hochschule der Bundeswehr, Munich.

LOHMÖLLER, J.-B. and WOLD H. (1980). *Three-mode path models with latent variables and PLS parameter estimation.* Forschungsbericht 80: 3, Fachbereich Pädagogik' Hochschule der Bundeswehr, Munich.

LYTTKENS, E., ARESKOUG, B. and WOLD, H. (1975). *The convergence of NIPALS estimation procedures for six path models with one or two latent variables.* Research Report 1975 : 3, Department of Statistics, University of Göteborg.

PIAGET, J. (1932). *The Moral Judgment of the Child.* Glencoe, Ill: 1948. First published in French.

STONE, M. (1974). *Cross-validatory choice and assessment of statistical predictions*, Journal of the Royal Statistical Society, Series B 38, 111-133.

STOUFFER, S.A. et al. (1949). *The American Soldier.* Studies in Social Psychology in World War II, Vols. 1 and 2. Princeton Univ. Press.

WOLD, H. (1966). *Nonlinear estimation by iterative least squares procedures.* Pp. 411-444 in Research Papers in Statistics, Festschrift for J. Neyman, ed. F.N. David, New York, Wiley.

WOLD, H. (ed. 1975). *Modeling in complex situations with soft information.* Group report at Third World Congress of Econometrics, 21-26 August, Toronto.

WOLD, H. (1977). *On the transition from pattern recognition to model building.* Pp. 536-549 in Mathematical Economics and Game Theory, Essays in Honor of Oskar Morgenstern, eds. R. Henn and O. Moeschlin, Berlin, Springer.

WOLD, H. (1979). *Model construction and evaluation when theoretical knowledge is scarce.* Cahier 79.06, Dept. of Econometrics, University of Geneva.

WOLD, H. (1980). *Soft modeling: intermediate between traditional model building and data analysis.* Banach Publications, Vol. 6, Mathematical Statistics, 333-346, Warsaw.

WOLD, H. (1982). *Soft modeling: the basic design, and some extension.* Chapter 1 in Wold (ed. 1982).

WOLD, H. (ed. 1982). *Systems Under Indirect Observations,* Part II. Amsterdam, North-Holland Publ.

RESUMÉ

Les idées développées dans ce texte s'inspirent de l'approche des modèles à variables latentes par les moindres carrés partiels (PLS) dite plus simplement « modélisation souple » (soft modeling), (Wold 1975, 1977, 1979, 1980, 1981).

Ce travail montre que la démarche suivie dans cette modélisation s'adapte facilement à l'analyse de variables dichotomiques formant une table de contingence multiple. Le modèle définit une (ou plusieurs) variable latente pour chacune des marges du tableau de contingence. Les variables observées s'interprètent comme des indicateurs dichotomiques de la variable latente correspondante et chaque variable latente sera d'ailleurs estimée comme un agrégat de ses indicateurs.

Le modèle est bâti sur des relations externes (outer relations) qui lient chaque variable latente à ses indicateurs ainsi que sur des relations internes (inner relations) qui mettent en rapport les variables latentes. Ces relations internes et externes jouent un rôle que l'on peut qualifier de causal-prédictif. En effet, dans le cas des indicateurs d'une variable latente elle-même expliquée par d'autres variables latentes, des substitutions simples permettent d'obtenir des relations liant les indicateurs aux variables latentes explicatives de la relation interne. La qualité prédictive du modèle peut être testée au moyen d'un test (Stone et Geisser) qui fournit des R^2 sans perte de degré de liberté.

Plusieurs généralisations se laissent envisager: effets de feed-back, structure hiérarchique des variables latentes ainsi que la multidimensionnalité de celles-ci.