I C U S S E M I N A R

Washington, September 3 - 5, 1984

....

RECENT DEVELOPMENTS ON CATEGORICAL DATA ANALYSIS BY PLS MODELING

by Jean-Luc Bertholet* and Herman Wold**

In the PLS (Partial Least Squares) approach to categorical data the observations are considered as manifest variables ruled by an unobservable latent model . This latent model is a system of exchange of information. The identification of the flows of information and their interpretation are the purposes of our paper.

Contents:

(*) University of Geneva, Department of Econometrics .
Sociological Research Service, Public Education Department, Geneva.

(**) Uppsala University.
University of Geneva, Department of Econometrics .

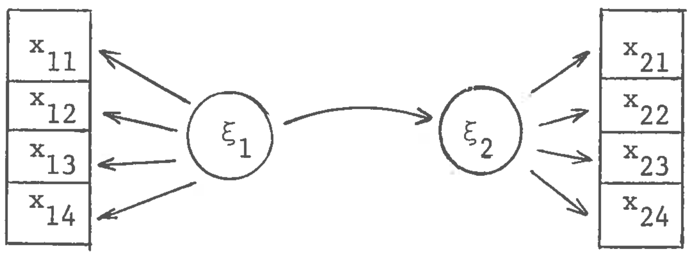
1 INTRODUCTION

1.1 Description of PLS

The partial least squares method (PLS) was introduced by H. Wold(*) to analyse data for which the theory was scarce: large data bank and primitive theory were the joint incentives which led to PLS. The PLS algorithm was primarily designed for scalar data. In 1981 some PLS analyses were applied on qualitative data, ie contingency tables (Wold and Bertholet 1981). Our first results seemed promising and invited to a deeper study. This paper presents some aspects of PLS analysis for qualitative data. Let us first give a briefing of PLS analysis. A comprehensive exposition is found in H. Wold (1982).

A PLS model involves manifest variables (ie observations) and latent variables (not observable). The manifest variables are grouped into blocks of indicators, one block for each latent variable. There is no restriction upon the number of indicators for a latent variable. The core of the model is a set of inner relations between the latent variables. These relations are illustrated by arrows that link the latent variables. The arrow scheme is the conceptual design of the model. Note that all information between the manifest variables is assumed to be conveyed by the latent variables. Here is for example a simple two-block model:

Figure 1. A two blocks model



Each block is here composed by four manifest variables ($x_{11} \ldots x_{14}$, $x_{21} \ldots x_{24}$). The arrows going from the manifest variables to the latent variables (ξ_1, ξ_2) form the outer model. The arrow between ξ_1 and ξ_2 is the inner model. In the basic PLS design the inner and outer relations are linear.

(*) For a complete exposition and references to earlier works (1973-77), see H. Wold (1982), Ch. 1.

Without loss of generality the following exposition disregards the location parameters; all variables are centered to zero mean.

Any PLS model is made up of four kinds of relations. Relations 1-2 constitute the theoretical model, relations 3-4 the estimated model. The fourth type of relation, as we shall see, can take two different forms:

Relations 1 The inner model is given by

$$\xi_2 = \rho_{12}\xi_1 + \varepsilon \tag{1.1}$$

and is subject to the predictor specification

$$E(\xi_2|\xi_1) = \rho_{12}\xi_1 .$$

The latent variables (LVs) are centered to zero mean

$$E(\xi_1)=E(\xi_2)=0 .$$

Relations 2 The outer relations link the manifest variables to the LVs of their block :

$$x_{1k} = \pi_{1k}\xi_1 + \nu_{1k} \tag{1.2}$$

with predictor specification :

$$E(x_{1k}|\xi_1) = \pi_{1k}\xi_1$$

x_{1k} in this section is the kth manifest variable in the first block (similarly for x_{2j} and ξ_2).

As π_{ik} and ξ_k are unknown, a standardization for model unambiguity is introduced:

$$E(\xi_1^2)=E(\xi_2^2)=1 .$$

The π_{ik} are called the loadings. The predictor specification implies:

$$E(\nu_{1k})=E(\nu_{2j})=E(\nu_{1k}\xi_1)=E(\nu_{2j}\xi_2)=0 .$$

PLS assumes that all information is conveyed by the LVs , hence:

$$E(\nu_{1k}\nu_{2j})=E(\nu_{1k}\xi_2)=E(\nu_{2j}\xi_1)=0 .$$

Relations 3 The estimated latent variables (X_1 for ξ_1 and X_2 for ξ_2) are linear aggregates of their indicators

$$X_1 = \sum_k w_{1k}x_{1k} \qquad X_2 = \sum_j w_{2j}x_{2j} \tag{1.3}$$

The w_{1k} and w_{2j} are called the weights of the indicators x_{1k} and x_{2j}.

Relations 4 In the estimation procedure the weights are estimated using the weight relations. These auxiliary relations can take two forms, called Mode A and Mode B.

Mode A

$$x_{1k} = w_{1k}X_2 + d_{1k} \quad (1.4a)$$

$$x_{2j} = w_{2j}X_1 + d_{2j}$$

Mode B

$$X_2 = \sum_k w_{1k}x_{1k} + d_1 \quad (1.4b)$$

$$X_1 = \sum_j w_{2j}x_{2j} + d_2$$

Mode A consists of a simple regression between each manifest variable and the LVs adjacent to the block it belongs. LVs which are directly connected by an inner arrow in the arrow scheme are called adjacent.

If there were more than one adjacent LV, for example a third LV connected to the first one, then the weight relation Mode A of the first block would be:

$$x_{1k} = w_{1k}(s_{12}X_2 + s_{13}X_3\ ) + d_{1k} \quad (1.5)$$

where s_{1m} = +1 or -1 according to the sign of the empirical correlation between X_1 and X_m.

In accordance with the standardization for model unambiguity the weights of block 1 are divided by a scalar g_1 so that X_1 have unit variance (similarly for X_2).

Mode B consists of a multiple regression where the regressors are the manifest variables and the regressand is the sign-weighted sum of the adjacent LVs of the block. If there were a third block connected to the first one, the Mode B weight relation of block 1 would be:

$$s_{12}X_2 + s_{13}X_3 = \sum_k w_{1k}x_{1k} + d_1 \ . \quad (1.6)$$

It is possible to mix in the same model Mode A and Mode B specifications. Nevertheless the following developments focus exclusively on the Mode B approach. A discussion of Mode A may be found in Wold and Bertholet (1981). Mode A is technically the same for scalar and categorical data whereas Mode B needs a small adaptation, see below section 2.3.2.

1.2 Estimation procedure

Before explaining the estimation procedure, we introduce some matrix notations we shall use.

The covariance matrix of the manifest variables is partitioned into blocks corresponding to the different blocks of the PLS model. For our two-block example, the partitioned covariance matrix of the manifest variables reads

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} ,$$

C_{ii} is called the <u>within</u> covariance matrix, C_{ij} $i \neq j$ is called the <u>between</u> covariance matrix.

The weights of a block are grouped into a column-vector w_i, the loadings are grouped into a column-vector π_i , and the estimated loadings are grouped into p_i. The dimension of these vectors is equal to the number of manifest variables in the ith block, say k_i.

The estimated LVs are grouped into vectors X_i with N row -entries where N is the number of observations.

The manifest variables of block i=1,2, after their means have been removed, are grouped into matrix Z_i with N rows and k_i columns. The covariance matrices are now, using (') to denote transpose :

$$C_{ii} = Z_i' Z_i (1/N) \qquad C_{ij} = Z_i' Z_j (1/N) \qquad (1.7)$$
$$i,j=1,2$$

The estimation process follows an algorithm based on ordinary least squares procedure that has two stages.

<u>Stage 1</u>

This stage alternates between (1.3) and (1.4) in order to get an estimation of the weights and the LVs. Here is the description of the step going from iteration s-1 to s:

The algorithm first computes the estimated LVs of iteration s using the weights of iteration s-1.

$$X_1^{(s)} = Z_1 w_1^{(s-1)} \qquad X_2^{(s)} = Z_2 w_2^{(s-1)} \quad . \qquad (1.8)$$

The dimensions of these vectors and matrices are:

$X_1(N*1)$, $Z_1(N*k_1)$, $w_1(k_1*1)$

$X_2(N*1)$, $Z_2(N*k_2)$, $w_2(k_2*1)$

$C_{12}(k_1*k_2)$, $C_{11}(k_1*k_1)$.

These LVs enter now the weight relations (1.4) whose ordinary least squares estimates are given in matrix notation:

<u>Mode A</u>

$$N^{-1} X_2'^{(s)} Z_1 = w_2'^{(s-1)} C_{21} = w_1'^{*(s)} \qquad (1.9a)$$

$$N^{-1} X_1'^{(s)} Z_2 = w_1'^{(s-1)} C_{12} = w_2'^{*(s)}$$

where the notation w_i^* will be explained in (1.10).

Mode B

$$X_2'^{(s)} Z_1 (Z_1' Z_1)^{-1} = w_2'^{(s-1)} C_{21}\ C_{11}^{-1} = w_1'^{*(s)} \qquad (1.9b)$$

$$X_1'^{(s)} Z_2 (Z_2' Z_2)^{-1} = w_1'^{(s-1)} C_{12}\ C_{22}^{-1} = w_2'^{*(s)}$$

The weights thus obtained must be rescaled in order to get LVs with unit variance, consequently the weights are divided by a scalar g_i:

$$w_i^{(s)} = w_i^{*(s)}\ (1/g_i^{(s)}) \qquad g_i^{(s)} = \sqrt{w_i'^{*(s)} C_{ii} w_i^{*(s)}} \qquad (1.10)$$

The weights enter now equations (1.8 a or b) to initiate a new iteration. This first stage of the algorithm stops when some conventional criterion is fulfilled. Each step s-1,s,s+1... uses ordinary least squares estimation. Experience shows that the method converges rapidly.

Stage 2

The weights and LVs estimated in the first stage enter now the relations (1.1) and (1.2) from which the loadings and the inner relations parameters are estimated,

$$\text{est}(\pi_i) = p_i = C_{ii}\ w_i \ . \qquad (1.11)$$

For the two-block model the path coefficient is the correlation between the estimated LVs,

$$\text{est}(\rho_{12}) = b_{12} = w_1'\ C_{12} w_2 \ . \qquad (1.12)$$

Comments: Inspection of the equations (1.8) and (1.9) shows that the ordinary least squares solutions of the weight relations Mode A and B can be written in a more similar and compact way :

mode A	Mode B	
$w_2'\ C_{21} = g_1\ w_1'$	$w_2'\ C_{21} = g_1\ p_1'$	(1.13) (a-b)
$w_1'\ C_{12} = g_2\ w_2'$	$w_1'\ C_{12} = g_2\ p_2'$	

and for both modes $p_i = C_{ii} w_i$.

CHAPTER 2. SOME REMARKS ABOUT PLS MODE B AND ITS APPLICATION TO CATEGORICAL DATA

Section 2.1 presents some properties of PLS mode B estimation, in particular its relation to canonical correlations analysis. Section 2.2 generalizes the propositions stated in section 2.1. Section 2.3 introduces some specific adaptations when dealing with categorical data. Section 2.4 is a discussion about latent variables (LVs) with more than one dimension.

2.1 The PLS Mode B approach and canonical correlations

We have seen in Chapter 1 (1.13b) that the ordinary least squares solutions of a two-block model are

$$w_2' C_{21} = g_1 \, p_1' \qquad\qquad p_1 = C_{11} w_1$$
$$w_1' C_{12} = g_2 \, p_2' \qquad\qquad p_2 = C_{22} w_2 \, . \qquad (1.13b) \text{ (again)}$$

These two equations can be joined to give

$$w_1' C_{12} C_{22}^{-1} C_{21} C_{11}^{-1} = g_1 \, g_2 \, w_1'$$

$$w_2' C_{21} C_{11}^{-1} C_{12} C_{22}^{-1} = g_1 \, g_2 \, w_2' \qquad (2.1)$$

which are the classical canonical correlations relations, Wold (1982). This is a well known result for two-block canonical correlations. An interesting problem is whether PLS can generalize canonical correlations to more than two blocks of variables. We shall see how PLS can solve this question.

Canonical correlations have several extensions to more than two sets of variables; Press (1972, p. 339) points out three extensions proposed by Anderson (1958), Horst (1965) and Kettenring (1969). The following section develops PLS in a way rather similar to Horst's proposition.

2.2 PLS modeling with more than two blocks

This section shows how PLS, when based on Mode B, gives an extension of canonical correlations.

We consider M sets of manifest variables (M>2). The classical approach of canonical correlations maximizes the correlation between two linear combinations of the variables of each set. With more than two blocks, we propose to look for the stationary point of the sum of the absolute value of the correlations between all the linear combinations of the sets taken two by two, under the constraint that the variance of each set is unity:

$$\sum_{i<j}^{M}\sum^{M} |w'_i\, C_{ij}\, w_j| \quad \text{under constraint } w'_i C_{ii} w_i = 1 \ , \ i=1\ldots M \qquad (2.2)$$

From now on we rewrite the absolute values as $s_{ij}w'_iC_{ij}w_g$, where s_{ij} is either +1 or -1. The Lagrangian is

$$L = \sum_{i<j}^{M}\sum^{M} s_{ij}w'_iC_{ij}w_j - (1/2)\sum_{i}^{M} q_i(w'_i\, C_{ii}w_i - 1) \qquad (2.3)$$

where q_i are the Lagrangian multipliers. The derivatives give:

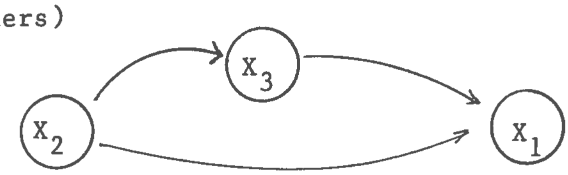
$$\sum_{t\neq i}^{M} s_{it}C_{it}w_t - q_iC_{ii}w_i = 0 \qquad (2.4)$$

$$w'_iC_{ii}w_i = 1 \ . \qquad (2.5)$$

The solution of these equations is not as simple as in the two blocks case because the Lagrangian multipliers q_i are all different. Nevertheless these two groups of equations (2.4) and (2.5) are very similar to the weight relations Mode B for a particular inner model, here called Complete Causal Chain. This model has the following stucture:

The first LV is the only exogenous LV, all subsequent LVs depend on this first. Considering M LVs, there will be M-1 arrows going from the first to the others. The second LV depends upon the first, and is directly influencing the M-2 subsequent LVs.

For a three LVs model, a Complete Causal Chain can be (among five others)



More generally, the matrix of the inner relations of a Complete Causal Chain is

$$\begin{pmatrix} 0 & & & & \ldots & 0 \\ 1 & 0 & & & \ldots & 0 \\ 1 & 1 & 0 & & & 0 \\ 1 & 1 & 1 & 0 & & 0 \\ & \ldots & & & & \\ 1 & 1 & 1 & \ldots & 1 & 0 \end{pmatrix} .$$

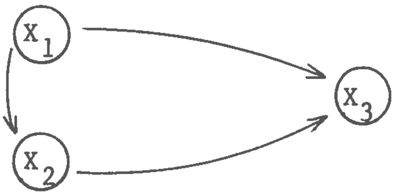
Equations (2.4) and (2.5) enter the PLS algorithm under the form:

$$\sum_{t\neq i}^{M} s_{it}\, C_{ii}^{-1} C_{it} w_t = w_i^* \tag{2.6}$$

$$w_i = w_i^* \sqrt{(1/\ w_i'^{*} C_{ii} w_i^*)} \tag{2.7}$$

The still unsolved question is about the convergence and uniqueness of the algorithm. Paul Horst (1965, p. 587) confesses in a related situation that no rigorous proof is known.

The previous discussion relates to a particular model (Complete Causal Chain). An important feature of PLS is to provide a variety of specifications for the inner relations. Let us now consider another model where only some of the correlations enter the summation (2.2). The author of the model then has built a path model in accordance with his assumptions; consequently the optimization involves only the covariances which occure in the inner model. Here are two illustrations with three-block models:



Complete Causal Chain,

$$\begin{bmatrix} \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot \\ 1 & 1 & \cdot \end{bmatrix}$$

inner model design matrix.

A specific model,

$$\begin{bmatrix} \cdot & \cdot & \cdot \\ 1 & \cdot & \cdot \\ \cdot & 1 & \cdot \end{bmatrix}$$

inner model design matrix.

2.3 Qualitative data and PLS mode B

2.3.1 Modeling of categorical data, PLS mode B

We present here the way PLS carries over to categorical data, ie contingency tables. Our working example will be a three-way contingency table with N observations (cases). Each of the N observations takes the form of three column-vectors x_{1n}, x_{2n}, x_{3n}. The dimensions of these vectors are the dimensions of the margins they refer to. All the entries of these vectors are zero except at the row corresponding to the characteristic of the nth observation. If the nth case shows the second characteristic in margin 1, then the transposed vector of this variable reads

$$x'_{1n} = (0\ 1\ 0\ 0 \ldots . 0)\ . \tag{2.8}$$

We group then these three transposed vectors in three matrices Z_1^*, Z_2^* and Z_3^* with N rows and k_1 ,k_2 and k_3 columns, where k_i is the number of categories in the ith margin.

We have $(1/N)Z'^*_i \iota_N = f_i$, where ι_N is a vector of ones and f_i the observed marginal frequency of the ith variable (ie the ith margin). Substracting from Z_i^* its mean we obtain the centered observations of the ith margin

$$Z_i = Z_i^* - \iota_N f'_i\ . \tag{2.9}$$

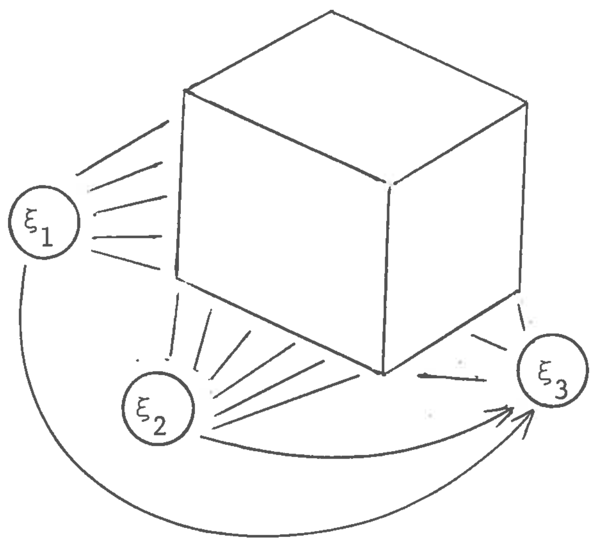
Here is briefly an example of a simple model for a three-way table illustrated by Figure 2.

A latent variable (X_1, X_2, X_3) is attached to each margin. All information between the margins is transmitted by these LVs; this information is conveyed according to a flow chart given by the arrows that link the LVs . The estimated inner relation is $X_3 = b_{13}X_1 + b_{23}X_3 + e_3$. The N case values of the LVs for the ith margin are estimated by

$$X_i = Z_i\, w_i\ , \tag{2.10}$$

where X_i is the vector of the N case values of the ith LV.

Figure 2. A PLS model applied to a three-way contingency table



Here, as always when each margin defines just one LV, the PLS model has an interesting feature: the input necessary for the estimation algorithm involves only information from the two-way marginal tables, ie the faces of the cube. The input covariances of the algorithm are

$$C_{ij} = Z_i' Z_j (1/N) = Z_i'^* Z_j^* (1/N) - f_i f_j' . \qquad (2.11)$$

More complex specifications can be developed which introduce interactions among variables; see Wold and Bertholet (1981) for an application with PLS Mode A.

2.3.2 PLS Mode B adaptation

Mode B is not directly applicable to qualitative data matrix because the within covariance matrices (C_{ii}) are not full rank: (2.11) multiplied by $\imath_{k_i}$ is zero. A unique solution for the loadings cannot be found, nevertheless each LV is uniquely determined as we shall prove it in this section.

The weight and loading relations for the three-block model are :

$$w_2' C_{21} = g_1 p_1' \qquad w_1' C_{12} = g_2 p_2' \qquad s_{13} w_1' C_{13} + s_{23} w_2' C_{23} = g_3 p_3' \qquad (2.12)$$

$$C_{11} w_1 = p_1 \qquad C_{22} w_2 = p_2 \qquad C_{33} w_3 = p_3 \qquad (2.13)$$

At each step of the algorithm, the relations (2.12) allow us to compute the loadings (p_i). From (2.13) one has to calculate the weights (w_i) in order to introduce them in the next step.

For categorical data the matrix C_{ii} cannot be inverted because its rank is k_i-1. But the rank of the bordered matrix of the linear systems (2.13) remains k_i-1, hence a set of solutions exists, we may choose one of them. It is easily verified that the rank of the bordered matrix is k_i-1 because $\imath_{k_i}' p_i = 0$ (postmultiplying (2.12) by $\imath_{k_i}$ and using $C_{ji} \imath_{k_i} = 0$ imply $\imath_{k_i}' p_i = 0$).

Recall that $C_{ii} = \mathrm{diag}(f_i) - f_i f_i'$ where f_i is the vector of frequencies of the ith margin of the table so that (2.13) reads

$$C_{ii} w_i = \mathrm{diag}(f_i) w_i - f_i (f_i' w_i) = p_i . \qquad (2.14)$$

The product $f_i' w_i$ is constant for each of the elements of p_i. Otherwise we see that the case values of the LVs computed by (2.10) do not depend on this constant, the LVs are thus uniquely determined. Taking this constant to be zero, the solution of (2.14) is immediate matter:

$$w_i = \mathrm{diag}(f_i)^{-1} p_i \qquad \text{and} \qquad C_{ii} w_i = p_i . \qquad (2.15)$$

The constraints on the parameters are now:

$$\imath'_{k_i} p_i = 0 \tag{2.16}$$

$$f'_i w_i = 0 \tag{2.17}$$

$$w'_i p_i = 1 \tag{2.18}$$

(2.16) is demonstrated in the previous paragraph. Multiplying (2.15a) by f_i' and using (2.16) proves (2.17). The standardization rule for unambiguity of the model is in matrix notation $w'_i C_{ii} w_i = 1$, use of (2.15b) proves (2.18).

Now, the case value of the LVs according to (1.8) and using (2.17) are given by the weights of the category of each observation. If, for example, the nth observation belongs to the second category in the ith margin, its case value LV is w_{i2}.

2.4 Latent variables with many dimensions

2.4.1 The second dimension of the between covariance matrices

Methods based on canonical correlations or principal components can proceed step by step in order to extract successive dimensions from the data. PLS can do it in a similar way. The dimension of the LVs and of their related parameters is now indicated by a superscript.

The outer relations with two dimensions in the LVs read

$$x_i = p_i^{(1)} X_i^{(1)} + p_i^{(2)} X_i^{(2)} + u_i^{(2)} \tag{2.19}$$

with

$$r(X_1^{(1)} X_2^{(1)}) = b_{12}^{(1)} , \qquad r(X_1^{(2)} X_2^{(2)}) = b_{12}^{(2)} ,$$

$$X_i^{(d)} = Z_i w_i^{(d)} .$$

We present here the algorithm proposed by H. Wold; it will be noted that J.-B. Lohmoeller introduced another approach based on simultaneous rotations (1981).

As usual when dealing with multidimensionality, it is assumed that the correlation between the LVs in a same block is zero:

$$r(X_i^{(d)} X_i^{(d')}) = 0 \qquad d \neq d' . \tag{2.20}$$

To fulfill this constraint, Wold uses the residuals of the outer relations of dimension 1 as input for the second PLS dimension

$$x_i = p_i^{(1)} X_i^{(1)} + u_i^{(1)} \Longrightarrow u_i^{(1)} = p_i^{(2)} X_i^{(2)} + u_i^{(2)} . \tag{2.21}$$

The covariance matrices are now

$$(I - p_i^{(1)} w_i'^{(1)}) C_{ij} (I - w_j^{(1)} p_j'^{(1)}) \qquad i \neq j . \tag{2.22}$$

In the process of estimation the within covariance matrices remain unchanged, and this gives interesting properties for the parameters (see below).

The third dimension is estimated similarly : the outer residuals of the second dimension become input for the third dimension.

2.4.2 Some comments

It can be easily demonstrated that the various dimensions of a LV are uncorrelated, but the dimensions of LVs belonging to different blocks are not expected to be uncorrelated. There are some particular situations where they are orthogonal, for example the two-block model leading to canonical correlations, or when there is a unique adjacent LV in the weight relation. It is too early in our research to undertake a general discussion about this important question.

The properties of the parameters and LVs with many dimensions are:

$$X_{\dot{i}}^{(d)} \text{ is orthogonal to } X_{\dot{i}}^{(d')} \text{ for } d \neq d', \qquad (2.23)$$

$$w_{\dot{i}}'^{(d)} \; p_{\dot{i}}^{(d')} = 1 \quad \text{if } d = d', \qquad (2.24)$$

$$w_{\dot{i}}'^{(d)} \; p_{\dot{i}}^{(d')} = 0 \quad \text{if } d \neq d'.$$

Demonstration of (2.23) is evident from ordinary least squares properties. Proposition (2.24) is interesting because it shows that the case values of the LVs of dimension larger than 1 can be computed with the original variables so that it is not necessary to use the residuals of the outer relations of the previous dimension. This appears clearly in equation (2.21a); multiplying it by the weights of the second dimension and using proposition (2.24) we infer:

$$w_{\dot{i}}'^{(2)} x_{\dot{i}} = w_{\dot{i}}'^{(2)} u_{\dot{i}}^{(1)} = X_{\dot{i}}^{(2)}.$$

This property is due to the specification of the between and within covariance matrices which enter PLS estimation Mode B.

2.4.3 Reconstruction of the data

After all dimensions have been extracted from the ith block, we can reconstruct the original variables by the use of the k_i-1 consecutive vectors of weights and loadings (k_i-1 is namely the rank of the within covariance matrix C_{ii}).

When a model involves blocks of unequal sizes, the question arises how to estimate more dimensions from the model than the smallest block allows to. This problem is discussed in Apel and Wold (1982) where a solution for PLS modeling is also given.

If all dimensions have been extracted and the parameters grouped into matrices denoted with capital letters, we obtain:

$$C_{ij} = Z_i'Z_j(1/N) = P_iW_i'Z_i'Z_jW_jP_j' = P_iR_{ij}P_j' , \qquad (2.25)$$

P_i is the matrix containing the consecutive estimated loadings of the ith block, W_i is the matrix of the successive estimated weights, R_{ij} is a matrix with k_i-1 rows and k_j-1 columns. This matrix contains the correlations between LVs of different dimensions of two blocks. As already mentioned it is diagonal when the jth block is the only adjacent block to the ith. We shall use this property in section 3.3.

CHAPTER 3. CONTINGENCY TABLE ANALYSIS : INTERPRETATION AND EXAMPLES

3.1 Why apply PLS to contingency tables

PLS modeling being distribution free, cf Wold (1982), it seems feasible to apply it to contingency table analysis. We are aware that many techniques and powerful tools have already developed in this area. We think PLS is a simple and informative method which brings new insight for prediction analysis.

In the PLS approach, the notion of prediction is understood as follows: The variable to predict, say x_2, is distributed in accordance to its marginal distribution, empirically f_2, so that any forecast of an outcome of x_2 will be based on f_2. If some extra information is given, for example the knowledge of an x_1 event, it is possible to improve the prediction by the use of a more accurate distribution for x_2. Then the marginal distribution of x_2 is no more optimal because, when x_1 has occured, x_2 is distributed in a different way. If x_1 brings no extra information -ie if the distribution of x_2 remains unchanged- x_1 is redundant and can be neglected.

To clarify this proposition we imagine first a two-way table where the cells are all independent, ie the frequencies are the products of the margins. Many models based on a chi-square or log-likelihood ratio will give very encouraging adjustment coefficients because few parameters are sufficient to describe entirely the cells of the table. In such a case PLS will be definitely discouraging because there is no predictive power in that table. No information about any variable can help to predict the other one more accuratly than its marginal distribution.

Let us now imagine another artificial example, a square table where all cells are zero except on the diagonal. Here again most of the current models will describe the entries of the table by the use of <u>complex</u> specifications (for example the presence of interaction parameters in a loglinear model, or the need for models based on eigenvalues decomposition to involve all the consecutive eigenvectors). In the PLS approach, we expect a very simple model to give the best prediction. Simplicity in PLS specification must originate from simplicity in the data, to summarize these last two paragraphs:

simple structure of the data -> simple PLS specification

no predictive power in the table -> bad adjustment coefficients.

Another important feature of PLS lies in the notion of causal chain. The contingency table is not considered as a set of numbers, but as an organized structure of data and flows of information. The inner model specification is the direct translation of the assumptions that the researcher introduces in his data.

3.2 Interpretation of the parameters

In PLS Mode B as defined in section 2.4 the weights and loadings are related by the relation $p_i = \mathrm{diag}(f_i)w_i$ so that the following discussion focuses essentially on loadings, correlations between LVs, and path coefficients of the inner model.

In this exposition, the manifest variables are in their original form, consequently the location parameters appear explicitly.

As in a previous paper (1982), we propose to interpret the parameters with reference to an approximate linear model.

The outer relations stated in relation (1.1) are

$$x_i = \pi_{0i} + \pi_{1i}\,\xi_i + \nu_i \qquad (3.1)$$

$x_i,\ \pi_{1i},\ \pi_{0i},\ \nu_i$ are k_i row vectors.

The x_i are distributed according to a multinomial distribution; hence

$$E(x_i) = \mathrm{prob}(x_i) = \pi_{0i} \quad . \qquad (3.2)$$

The estimated location parameters are the marginal distribution of the variables. Now by analogy to (3.2) we define

$$E(x_i|\xi_i) = \pi_{0i} + \pi_{1i}\,\xi_i = \mathrm{prob}(x_i|\xi_i) \quad . \qquad (3.3)$$

In the sample procedure the distribution of x_i changes from case to case around a mean value (π_{0i}) which is its marginal distribution.

When the correlation between ξ_i and another LV, say ξ_j , is high, the two corresponding margins are correlated. This correlation coefficient is an indicator of <u>dependence</u> between the two margins. It is a <u>global</u> measure between margins, whereas the loadings are related to the different <u>categories</u> which compose a margin.

The relation (3.3) shows explicitly that the higher the absolute value of a loading, the more is the corresponding category influenced by its LV. As an extreme case, a zero coefficient in a loading vector means that the category has the same relative frequency in the margin as within the table. We propose here an example of a 4*3 contingency table:

Table 1. artificial two-way contingency table

25 (0)	125 (0)	100 (0)	250
1 (-)	100 (-)	149 (+)	250
70 (+)	150 (+)	30 (-)	250
4 (-)	125 (0)	121 (+)	250
150	500	400	1000

Inner model: $X_{2n} = b_{12n} X_{1n} + e_{2n}$.

The estimated outer relations are:

$$x_{1n} = \begin{pmatrix} .25 \\ .25 \\ .25 \\ .25 \end{pmatrix} + \begin{pmatrix} .00 \\ -.24 \\ .40 \\ -.16 \end{pmatrix} X_{1n} + u_{1n} \qquad x_{2n} = \begin{pmatrix} .1 \\ .5 \\ .4 \end{pmatrix} + \begin{pmatrix} .24 \\ .14 \\ -.38 \end{pmatrix} X_{2n} + u_{2n}$$

The inner relation : $X_{2n} = .45 X_{1n} + e_{2n}$

In this example we do not report on the LVs and their case values. The estimated correlation between LVs is 0.45, the table is not independent when considered as a whole, note that the Cramer coefficient of association for this table is .26 . Nevertheless the first row is independent of x_2. The probability of an outcome of the first category of x_1 -we write here x_{11} - is not influenced by the outcome in the second margin. As the LVs are supposed to convey all the information between the margins, the estimated outer relation of the first category of x_1 reads:

$$\text{prob}(x_{11} \mid X_1) = E(x_{11} \mid X_1) = f_{11} + 0\, X_1 \quad .$$

The signs of the loadings help to interpret the entries of the table. Comparing the three columns of the table with their expected values under the independence assumption, we write (+) if the observed frequency is larger than the frequency under this independence assumption, (-) if it is smaller, and (0) if it is equal.

The two first columns behave in a very similar way, but the last one is just opposite in sign. The estimated loadings for x_2 reflect this feature: both loadings of columns 1 and 2 are positive (.24,.14) whereas the third is negative (-.38).

The sections 3.3-3.5 present numerical applications of PLS which will illustrate the above propositions of interpretation. Section 3.3 and 3.4 use artificial data, Section 3.5 analyses real-world data which have been analysed by other authors.

3.3 Partial independence

In this section we apply the PLS algorithm to contingency tables with a specific structure, namely partial independence. Two variables are said to be partially independent if, under the control of a third one, they become independent. This conceptual design is similar to the zero partial correlation of scalar variables.

If x_2 and x_3 are partially independent with respect to x_1, then

$$\text{prob}(x_2 x_3 | x_1) = \text{prob}(x_2 | x_1)\text{prob}(x_3 | x_1) \qquad (3.5)$$

but $\text{prob}(x_2 x_3) \neq \text{prob}(x_2)\text{prob}(x_3)$.

Under this condition the covariances matrices are related,

$$C_{21}\ \text{diag}(f_1)^{-1}\ C_{13} = C_{23} \quad . \qquad (3.6)$$

The interpretation of the partial independence in terms of prediction states that if one tries to predict x_3 when x_1 is known, then the extra information of x_2 is null, ie the variable x_2 does not play any role for predicting x_3 when x_1 is given. Note that x_2 and x_3 can be reversed in this statement.

Using this assumption we formulate the inner relation :

that is $X_3 = b_{13}X_1 + b_{23}X_2 + e_3$. (3.7)

Estimation of contingency tables which obey relation (3.6) give automatically the path coefficient $b_{23} = 0$ as expected. This path coefficient results from an ordinary least squares estimation of (3.7), to specify,

$$b_{23} = (r_{23} - r_{12}r_{13})/(1 - r_{12}^2) = 0 \qquad (3.8)$$

where r_{ij} is the correlation of X_i and X_j.

The expression (3.8) is interesting because it bridges scalar and categorical modeling. The partial correlation between the LVs X_2 and X_3 after removing the influence of X_1 is

$$\text{corr}(X_2 - r_{12}X_1)(X_3 - r_{13}X_1) \qquad (3.9)$$

$$= (r_{23} - r_{12}r_{13})/\sqrt{(1 - r_{12}^2)(1 - r_{13}^2)}$$

$$= b_{12}\sqrt{(1 - r_{12}^2)/(1 - r_{13}^2)} \quad .$$

The partial correlation coefficient is zero at the same time as the path coefficient. When partial independence holds for qualitative data, the estimated LVs, which are scalar variables, behave in the same way: their partial correlation is zero. There is a parallelism between scalar and categorical data.

Until now we have not demonstrated that assumption (3.6) implies $b_{23}= 0$. The demonstration involves the reconstruction of the data of section 2.4.3. The relation (3.5) then takes the form

$$P_2R_{21}P_1' \operatorname{diag}(f_1)^{-1} P_1R_{13}P_3' = P_2R_{23}P_3'$$

where $P_1'\operatorname{diag}(f_1)^{-1}P_1= W_1'P_1= I_{k_1-1}$

which gives:

$$P_2R_{21}R_{13}P_3' = P_2R_{23}P_3' \quad . \qquad (3.10)$$

A sufficient condition for (3.10) to hold is

$$R_{21}R_{13} = R_{23} \quad .$$

The diagonal elements of R_{ij} are the correlations between consecutive dimensions of the LVs. Now, the chosen specification of the inner model leads to

$$r(\ X_i^{(d)}\ X_j^{(d)}\ ) \neq 0 \qquad d = 1 \dots \min(k_i,k_j)$$

$$r(\ X_i^{(d)}\ X_j^{(d')}\ ) = 0\ , \qquad d \neq d' \ .$$

The R_{ij} matrices can be partitioned in a square diagonal matrix containing all correlations of the LVs that are different from zero plus a null matrix. Hence for $k_i < k_j$ we have

$$R_{ij} = \left[\operatorname{diag}(r(X_i^{(d)}\quad X_j^{(d)}\ ))\ \vdots\ 0 \right] \qquad d=1\dots k_i-1$$

where the order of R_{ij} is $(k_i-1)*(k_j-1)$.

Our sufficient condition is now :

$$r_{12}^{(d)}r_{13}^{(d)} = r_{23}^{(d)} \quad . \qquad (3.11)$$

Here is a numerical illustration for an artificial table with four margins and 3*4*2*4 cells. The marginal distributions are :

f_1' = (.1700 .3000 .5300)

f_2' = (.1743 .1209 .3466 .3582)

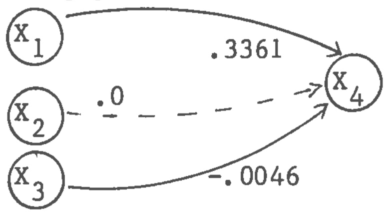
f_3' = (.9552 .0448)

f_4' = (.1360 .1984 .2672 .3984)

The partial independence was imposed on the variables 2 and 4 with respect to 1, that is :

$$\text{prob}(x_2 x_4 | x_1) = \text{prob}(x_2 | x_1)\text{prob}(x_4 | x_1) \quad .$$

The following path coefficients were estimated:



The first variable is the most important for predicting the fourth. The second does not bring any information, whereas the third has a small importance.

3.4 Simpson's paradox

This section shows a simple example of what is called Simpson's paradox. Here again artificial data describe a three-variable situation, namely sex (Male, Female), a medical treatment (T=yes, NT=no treatment), and the result of the treatment (R=succes, NR=failure). The odds ratios for the two populations of males and females show undoubtedly that the treatment has a negative effect, but once the two populations have been joined, the apparent effect is just the opposite. This example is inspired by Upton (1978, p. 43).

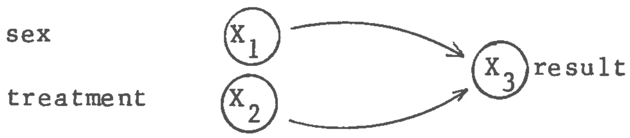
Table 2. Artificial three-way contingency table

M	T	NT
R	10	100
NR	100	500

F	T	NT
R	100	60
NR	50	20

F+M	T	NT
R	110	160
NR	150	520

Inner model:



The inner relation reads:

$$X_3 = b_{13}X_1 + b_{23}X_2 + e_3 . \tag{3.12}$$

The outer relation of "result" including location parameters reads:

$$x_3 = f_3 + p_3 X_3 + u_3 . \tag{3.13}$$

Substituting (3.12) in (3.13), we obtain the so called substitutive predictive relation, cf Wold (1982):

$$x_3 = f_3 + p_3(b_{13}X_1 + b_{23}X_3) + p_3 e_3 + u_3 \quad . \qquad (3.13)$$

Taking the deterministic part of (3.13), we define by analogy to (3.3):

$$\text{prob}(x_3|X_1X_2) = E(x_3|X_1X_2) = f_3 + p_3(b_{13}X_1 + b_{23}X_2) \qquad (3.14)$$

$$= \begin{pmatrix} .29 \\ .71 \end{pmatrix} + \begin{pmatrix} -.45 \\ .45 \end{pmatrix} (-.55X_1 + .08X_2)$$

The estimated LV for sex and treatment are:

sex: X_1	treatment: X_2
M : -.57	T : 1.62
F : 1.76	NT : -.62

Substitution of the LVs values for treatment and sex in (3.14) shows that treatment has a negative effect because it lowers the probability of success. The influence of sex is more important than treatment; compare .55 against .08. Otherwise females are more successful than males.

Note that, for this specification, the only data the algorithm needs are the two marginal tables sex*result and treatment*result. The joint distribution of the three variables does not participate the estimation. This example shows that PLS is able to give accurate results even with incomplete data sets. This is a very important feature because it often happens in practice that the only information consists of two-way contingency tables, the joint distribution being unknown.

3.5 Real-world data example

This last section presents real-world data quoted from Goodman (1973); they have been analysed by Lazarsfeld (1948,1968) and Lipset et al. (1954).

Reproduced in the Table 3, these data cross-classify responses of 266 people. Each person was interviewed at two successive moments. The questions were similar at both times: 1) vote intention (replublican/ not republican) 2) opinion of a particular candidate (for/against a republican candidate). The answers form a four-way contingency table. The succession of the interviews in time gives a general framework to the structure of the path models one can apply to these data.

Goodman proposes to develop the logarithm of the odds ratio as a sum of parameters which are directly related to those of the usual loglinear model. These parameters are selected in accordance to one of the 30 models that could describe the data. These parameters play the role of path coefficients. Two models which fit the data very well are reproduced below.

Table 3 Observed cross-classification of 266 people interviewed at two successive points in time.

		second interview			
	vote intention V_2	+	+	-	-
	candidate opinion C_2	+	-	+	-
first interview					
vote intention V_1	candidate opinion C_1				
+	+	129	3	1	2
+	-	11	23	0	1
-	+	1	0	12	11
-	-	1	1	2	6

Figure 3 Path diagram of model H5, which adjusts the margins (AB),(AC),(BC),(BD),(CD). Likelihood ratio = 1.46 df = 6. Goodman (1973/1978, p. 190)

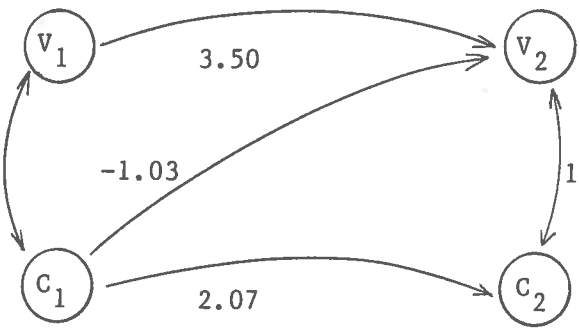
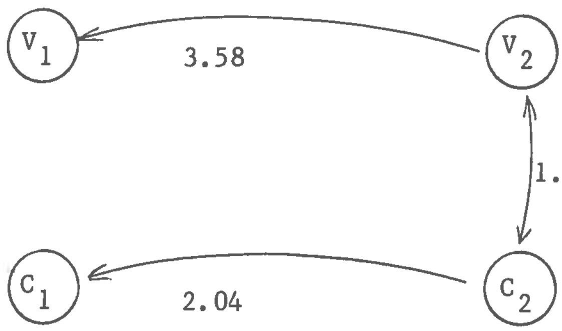
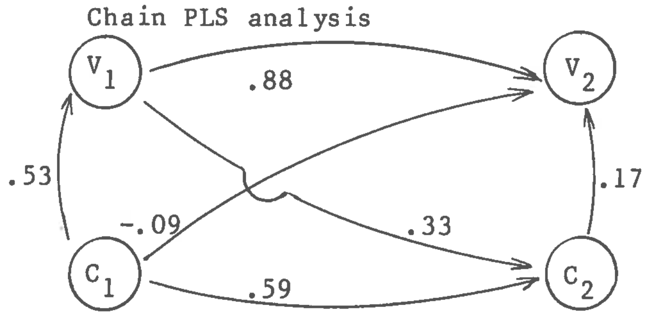


Figure 4 Path diagram of model H14 which adjusts the margins (AC),(BD),(CD). Likelihood ratio = 8.62, df = 8. Goodman (1973/1978, p. 195)



We shall now compare PLS models to the above results. We shall discuss three PLS models. The first is the Complete Causal Chain with the opinion of the candidate in the first interview (C_1) as exogenous LV. The two other models are similar to those reproduced from Goodman's article. Note that PLS does not have arrows pointed in both directions.

Figure 5 Path coefficients of a Complete Causal Chain PLS analysis



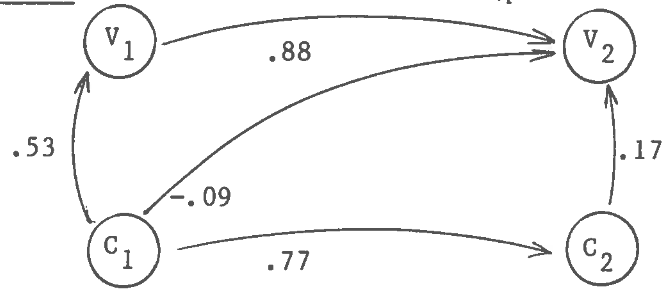
Multiple correlation coefficients (R^2) of the LVs corresponding to margin:

V_1	C_2	V_2
.28	.67	.90

Interpretation: The voting attitude at the second moment is mainly influenced by the voting attitude at the first time (V_1) and by the candidate opinion at the second interview (C_2). It seems that there is no direct memory of the opinion in the first interview; the influence of C_1 is transmitted by V_1 and C_2.

This general model is useful as a first approach to the data because it helps to develop more simple models by removing the less predominant relations.

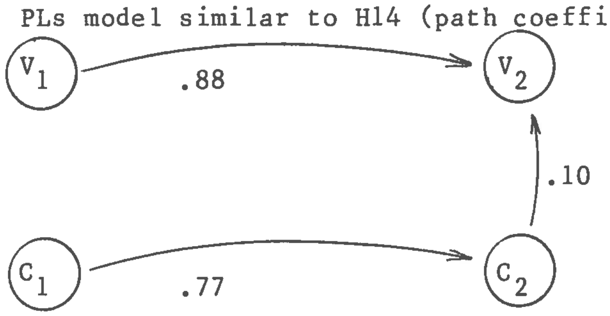
Figure 6 PLS model similar to H5 (path coefficients)



Multiple correlation coefficients (R^2) of the LVs corresponding to margins:

V_1	C_2	V_2
.28	.59	.90

Figure 7 PLs model similar to H14 (path coefficients)



Multiple correlation coefficients (R^2) of the LVs corresponding to margin:

C_2	V_2
.59	.90

The ranking order of the parameters is much the same in both loglinear H14 and PLS modeling, the only difference occures in the model H5 for the path coefficients between C_1-V_1 and C_2-V_2.

The PLS path coefficients are factor loadings of LVs with unit variance and zero mean, hence their sizes are directly comparable. In Goodman's approach the path parameters are logarithms that enter an addition equation, their range is much larger than in PLS.

Goodman's path model has been developed for dichotomous variables, hence the generalization of his method based on odds ratios is not straighforward. In PLS approach, the number of categories in a margin does not interfere at all with path modeling technique.

CHAPTER 4. JACKKNIFE TECHNIQUES

4.1. Jackknife techniques for contingency tables

Strictly speaking, jackknife methods should reestimate the model for each of the N observations. This may be time-consuming, so that simplified methods based on grouping data are often used.

For contingency table the jackknife may be applied in a very economic way. Let us take a simple example, a two by two table:

$$T = \begin{vmatrix} 20 & 12 \\ 11 & 10 \end{vmatrix} \qquad N=53 \quad .$$

The removal of one observation gives one of the four following tables

$$\begin{vmatrix} 19 & 12 \\ 11 & 10 \end{vmatrix} \begin{vmatrix} 20 & 11 \\ 11 & 10 \end{vmatrix} \begin{vmatrix} 20 & 12 \\ 10 & 10 \end{vmatrix} \begin{vmatrix} 20 & 12 \\ 11 & 9 \end{vmatrix} \quad .$$

Because of this automatic grouping the number of estimations needed for jackknife techniques is equal to the number of cells and not to the number of observations. This is important, for in practical situations there are often many more observations than cells.

4.2 Jackknife estimates of standard errors

We have applied the technique of the previous section to the real world data cross-classification by Lazarsfeld as modeled by our Complete Causal Chain in section 3.5.

The formulas are, using Efron and Stein's notation (1981):

$$est(var(S)) = (N-1)/N \sum_i (S(i) - S(.))^2 \qquad (4.1)$$

$$S(.) = (1/N) \sum_i S(i) \quad . \qquad (4.2)$$

S is a statistic invariant under any permutation of its arguments. S(i) is the value of that statistic when the ith observation has been deleted from the sample.

Under the classical - often unrealistic - assumptions of independence and identical distribution, Efron and Stein show that the jackknife estimate of variance is positively biased, its expectation being larger than the true variance of the statistic S.

The jackknife estimates of the Complete Causal Chain (cf figure 5) are given below:

parameters	estimates	jackknife variance estimates
b_{13}	.8813	.0022
b_{14}	.3336	.0042
b_{21}	.5329	.0029
b_{23}	-.0906	.0027
b_{24}	.5885	.0042
b_{43}	.1700	.0058

4.3 Stone-Geisser testing for predictive relevance

Stone (1974, p. 121) develops an interesting property of his prediction approach which we shall use here because there is no need of reestimation of the parameters. In Stone's notation let the prediction function be

$$\text{pred}(y) = \sum_k a_k b_k c_k(x) \quad , \tag{4.3}$$

where $c_k(x)$ are specified functions, a_k are 0 or 1 corresponding to a "choice of variables", and b_k are parameters to be estimated.

This formulation is useful for exploring a model by the use of the substitutive predictive relations; cf (3.13). Again we shall illustrate by our model for the real-world data of Lazarsfeld.

The endogenous variable to be predicted is the voting attitude at the second interview V_2 (with LV X_3) by the use of C_1,V_1 and C_2 (whose LVs are respectively X_2,X_1 and X_4). We have seen that the path coefficient (b_{23}) that links C_1 to V_2 is close to zero so that we shall explore if it is meaningful to omit C_1 in the prediction. We shall test if this simplification lowers the predictive power of the model, that is

$$\text{pred}(X_3) = b_{13}X_1 + b_{23}X_2 + b_{43}X_4 \tag{4.4}$$

against

$$\text{pred}(X_3) = b^*_{13}X_1 + b^*_{43}X_4 \ . \tag{4.5}$$

The parallels with Stone's relation (4.3) are

$$c_1(x) = X_1 \qquad c_2(x) = X_2 \qquad c_3(x) = X_3$$
$$b_1 = b_{13} \qquad b_2 = b_{23} \qquad b_3 = b_{43} \ ,$$

and we test a_1=a_2=a_3=1 against a_1=a_3=1.

In this test procedure the functions $c_k(x)$ are given; ie the weights and LVs are not reestimated. To repeat, the purpose of this test is to explore a complex specification to see if some relations are candidates for simplification. Note that, regarding (4.4) and (4.5) as different models, X_1 and X_2 should not be the same in both cases. If they were reestimated for (4.5), then the difference between the two loss functions would be smaller than the actual one.

Stone's loss function is

$$L(a) = (1/N) \sum_i (y_i - \text{pred}(y_i,a,S(i))^2 \ , \tag{4.6}$$

where pred(y_i,a,S(i)) is the predicted value for y_i, using specification "a", when the ith observation has been deleted from the sample during the estimation.

Stone shows that the general loss function (4.6) for the linear form (4.3) may be calculated in a simple manner:

$$L(a) = (1/N) \sum_i \left\{(y_i - y_i(a)) / (1-A_{ii}(a))\right\}^2 \qquad (4.7)$$

where $y_i(a)$ is, for fixed a_1, a_2 and a_3, the least squares prediction without any deletion in the sample. $A_{ii}(a)$ is the diagonal element of

$$A(a) = X(a)\left[X(a)'X(a)\right]^{-1} X(a)' \qquad (4.8)$$

which is the projection matrix of the least squares estimation of (4.4) or (4.5). In the jackknife estimation (4.7) there is no extra estimation except the matrix inversion in (4.8).

Stone-Geisser's test criterion, denoted Q^2, is an R^2 evaluated without loss of degrees of freedom, cf Wold (1982). It will be noted that

$$Q^2 = 1 - L(a) .$$

The results calculated for the models (4.4) and (4.5) are very similar:

model (4.4): $L(a) = .1065$, $Q^2 = .8935$

model (4.5): $L(a) = .1079$, $Q^2 = .8921$.

The difference is very small, showing that the PLS model maintains nearly the same predictive relevance if the relation $C_1 \rightarrow V_2$ is omitted.

CHAPTER 5

STRUCTURE ANALYSIS AND PLS

5.1 Introduction

This Chapter presents a recent field for PLS applied to qualitative data. The four previous chapters refer essentially to prediction and dependence. Here we shall focus on the pattern of a contingency table as modelled by a Guttman scale.

5.2 Guttman scales

Guttman scales are of frequent use in psychology and education. For example, a sample of school children are asked two questions (often called items). The answers to these questions take only two forms yes/no or success/failure, usually denoted by the codes 1 or 0.

The psychological or pedagogical theory often supposes a pattern for these answers. For example it may be expected that children who are able to solve a multiplication problem must also be able to solve an addition problem. This intuitive simple theory leads to the scheme:

$$\text{multiplication} \Longrightarrow \text{addition}$$

Conversely it may happen that children can solve the addition item but fail in the multiplicative one. Denoting by 1 the success and by 0 the failure, the theory expects the following pattern:

addition	multiplication
0	0
1	0
1	1

The answer <0 1> is in disaccord with our above assumed intuitive theory and therefore is not included in the set of expected answers.

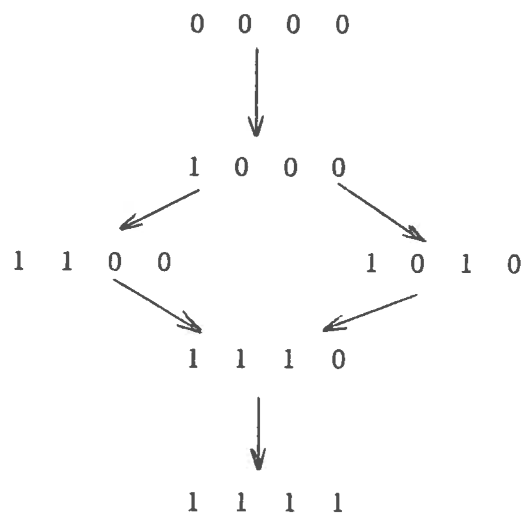
A Guttman scale is the hierarchic organization of such items. The scale may be generalized to more items; here is a four-item scale:

(5.1)

item 1	item 2	item 3	item 4
0	0	0	0
1	0	0	0
1	1	0	0
1	1	1	0
1	1	1	1

In this stringent pattern there is a perfect hierarchic organization of the items. It may happen that two items are considered as equivalent in the scale: only one or the other, or even both, can be successful without changing the order of the others. The pattern is then as follows:

(5.2)



This situation is called a biform scale. More complex patterns can be imagined: see H. Jörg Henning (1981).

Experimental data will often deviate more or less from the theoretical pattern and the question one has to answer is "to which extent do the experimental data support the structure assumed by the theory ?" . It is of course expected that some observations won't agree with the scale, but their frequency must be small enough for the assumed theory to be still relevant.

We shall present here a way PLS can analyse the experimental data for uniform scales. Our example is a four-item Guttman scale in the uniform pattern (5.1). Each item is now considered as a margin of a four-way contingency table with 2x2x2x2 = 16 cells. According to the theory some of these cells are empty. Using the scale given in (5.1) we shall consider the following theoretical contingency table in which the non empty cells are crossed (X).

(5.3)

			item 4	
item 1	item 2	item 3	0	1
0	0	0	X	.
0	0	1	.	.
0	1	0	.	.
0	1	1	.	.
1	0	0	X	.
1	0	1	.	.
1	1	0	X	.
1	1	1	X	X

5.3 PLS and uniform Guttman scale

Using a probability approach the uniform scale can be described by a set of conditional probabilities, namely:

$$\text{prob}(x_i = 1 \quad x_j = 1) = 1 \qquad \text{if } i<j \qquad i,j = 1,4. \qquad (5.4)$$

A smaller but equivalent set of conditions is

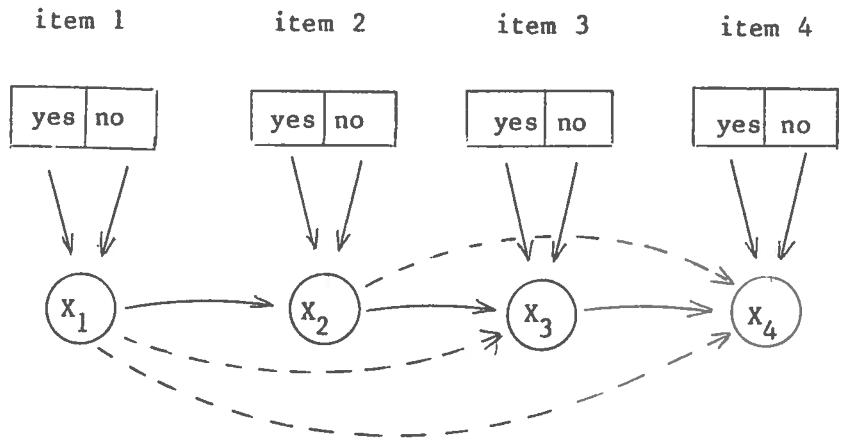
$$\text{prob}(x_i = 1 \quad x_j = 1) = 1 \qquad \text{if } j = i + 1 \qquad i = 1,2,3 \qquad (5.5)$$

where x_i denotes item i.

In Chapter 3 we have noted the similarity between prediction and conditional probability. As PLS is a predictive approach it seems feasible to apply it to the analysis of Guttman scales.

As we have four items, we shall build four corresponding LVs. Their indicators are the two answers (yes/no). The PLS inner relations will be the Complete Causal Chain, see 2.1.

(5.6)



It will be demonstrated that the PLS estimation of this inner specification on the theoretical table (5.3) gives the following results:

the arrows which do not link two consecutive LVs disappear, ie the corresponding path parameters become zero.

The starting form of the inner relations to be estimated read:

$$X_2 = b_{12} X_1 + u_2$$

$$X_3 = b_{13} X_1 + b_{23} X_2 + u_3$$

$$X_4 = b_{14} X_1 + b_{24} X_2 + b_{34} X_3 + u_4 \quad .$$

Then the results of the estimation are:

$$X_2 = b_{12} X_1 + u_2$$

$$X_3 = b_{23} X_2 + u_3 \qquad b_{13} = b_{14} = b_{24} = 0 \quad . \qquad (5.7)$$

$$X_4 = b_{34} X_3 + u_4$$

Note that the direction of the arrows can be inverted without changing the argument; the estimated inner relations in this case take the following form:

$$X_3 = b_{43} X_4 + u_3 \qquad X_2 = b_{32} X_3 + u_2 \qquad X_1 = b_{21} X_2 + u_1 \quad (5.8)$$

The demonstration is given in the annex.

We know that some path parameters are expected to be zero when the theoretical model of the uniform scale holds. If experimental data analyzed by PLS give the non-consecutive path parameters sufficiently close to zero, then one can accept the theoretical model. As we have not postulated any distributional form for the population, it is not possible to apply any classical statistical test procedure. Instead we may use the jackknife standard errors of the parameters as empirical criterion. In the next section we present three numerical examples as illustration of the sensitivity of the analyses.

5.4 Numerical examples of Guttman scale

On the next page is a table showing three sets of artificial data for a four-item uniform Guttman scale. The same inner model (5.7) is used in the three situations. The bottom part of the table reports the path parameters. The first example uses the theoretical structure of a uniform Guttman scale, the non empty cells being the same as in (5.3). In the second example some "unexpected" observations have been introduced into the cells that should be empty according to the assumptions. In the third example the cells corresponding to the theoretical model have been multiplied by ten in order to increase the discrepancy between expected and unexpected observations.

What about more complex scales? - We have actually no general method for the way of studying complex scales. Sometimes it may be feasible to mix two items of the original structure to obtain a unique item which takes information from its two constituents. Going back to the biform scale (5.2) we see that the items 2 and 3 are "equivalent", ie they both lead to the same structure <1 1 1 0> following alternative ways <1 0 1 0> or <1 1 0 0>. In this case, these two items could be combined by the logical "or" without changing the general structure of the scale. Our proposition is then to build a new item, say item 23, which joins items 2 and 3 according to the following definition :

item 23 = item 2 <u>or</u> item 3

This change leads us to a three-item uniform case which again is easily analyzed by PLS.

<u>Uniform scale examples</u>

Items	example 1		example 2		example 3	
1 2 3 4:	0	1	0	1	0	1
0 0 0	10	.	10	1	100	1
0 0 1	.	.	.	2	.	2
0 1 0	.	.	1	1	1	1
0 1 1	.	.	1	.	1	.
1 0 0	25	.	25	3	250	3
1 0 1	.	.	2	.	2	.
1 1 0	18	.	18	2	180	2
1 1 1	26	78	26	78	260	780
b_{12}	-.487		-.419 (.082)		-.477 (.023)	
b_{13}	.000		.035 (.080)		.005 (.011)	
b_{14}	.000		.033 (.075)		.006 (.011)	
b_{23}	-.750		-.651 (.068)		-.738 (.018)	
b_{24}	.000		-.081 (.092)		.013 (.014)	
b_{34}	-.709		-.550 (.084)		-.689 (.018)	

The standard errors have been computed using the jackknife technique of Chapter 4.

In the artificial table with the theoretical structure (5.3) the path parameters which do not link two consecutive LVs are zero. In the second example thirteen observations have been put into empty cells of the previous table. Now all path parameters are different from zero. The proportion of these "unexpected" observations amounts to 7.6% of the total number of cases. In the last example this percentage amounts to 0.8; as it is weaker, the estimated parameters are much closer to the theoretical values.

ANNEX : UNIFORM GUTTMAN SCALE

The answers to four items are of the 0/1 form. We postulate that the pattern of the answers is given by the following set of answers:

item 1	item 2	item 3	item 4
0	0	0	0
1	0	0	0
1	1	0	0
1	1	1	0
1	1	1	1

(C.1)

We shall demonstrate that when the model (C.2) is estimated by PLS for a sample of answers that belong to the pattern (C.1), then $b_{ij} = 0$ if $j \neq i + 1$.

$$\begin{aligned} X_2 &= b_{12} X_1 + u_2 \\ X_3 &= b_{13} X_1 + b_{23} X_2 + u_3 \\ X_4 &= b_{14} X_1 + b_{24} X_2 + b_{34} X_3 + u_4 \end{aligned} \qquad (C.2)$$

To show this, we shall use the inner relations in their OLS estimated form. Let us write the covariance matrix of the estimated LVs and then the solution of normal equation for the last inner relation (X_4),

$$\begin{matrix} X_1 : \\ X_2 : \\ X_3 : \\ X_4 : \end{matrix} \left(\begin{array}{ccc|c} 1 & r_{12} & r_{13} & r_{14} \\ r_{21} & 1 & r_{23} & r_{24} \\ r_{31} & r_{32} & 1 & r_{34} \\ \hline r_{41} & r_{42} & r_{43} & 1 \end{array} \right) = \left(\begin{array}{c|c} R_{xx} & R_{xy} \\ \hline R_{yx} & R_{yy} \end{array} \right) . \qquad (C.3)$$

Hence:

$$\begin{pmatrix} b_{14} \\ b_{24} \\ b_{34} \end{pmatrix} = R_{xx}^{-1} R_{xy} = \begin{pmatrix} 0 \\ 0 \\ b_{34} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} b_{34} \qquad (C.4)$$

where the first term is the general solution and the two last hold for the uniform Guttman scale. To demonstrate this property, we note that in the last term R_{xy} is proportional to the third column of a matrix for which R_{xx}^{-1} is the inverse. This column is then the third column of R_{xx}. To prove our theorem, it is sufficient to show :

R_{xy} is proportional to the last column of R_{xx}. (C.5)

This statement does not depend on the number of items; the following demonstration is general.

We know that the elements of R_{xy} and R_{xx} are correlation coefficients given by:

$$r_{ij} = \underline{w}_i' C_{ij} \underline{w}_j \qquad (C.6)$$

In this annex we underline the vectors, that is $\underline{w}_i' = [w_{io} \; w_{i1}]$

The subscript io means "0" type of answer to the i^{th} item, whereas i1 means "1" type of answer.

We write the condition (C.5) as:

$$r_{ij} = q \, r_{i\;j+1} \quad , \qquad (C.7)$$

where $q \neq 0$ and i smaller than the index of the LV explained by the inner relation; in our example (C.4) $i<4$. By (C.6) this condition reads:

$$\underline{w}_i' C_{ij} \underline{w}_j = q \, \underline{w}_i' C_{i\;j+1} \underline{w}_{j+1}$$

We have then to study the properties of C_{ij}. As the weights are centered to zero with respect to the marginal distributions, we may work with the product matrices instead of the covariance matrices, that is:

$$\underline{w}_i' C_{ij} \underline{w}_j = \underline{w}_i' \left[F_{ij} - f_i f_j'\right] \underline{w}_j = \underline{w}_i' F_{ij} \underline{w}_j \qquad (C.8)$$

F_{ij} is the cross distribution of items i and j; f_i is the marginal distribution of item i.

Let us now introduce some notations. For $j>i$ the pattern of answers allowed by the assumptions (C.1) is:

item i	item j	frequency
0	0	$f_{ij}^{oo} = f_i^o$
1	0	f_{ij}^{1o}
1	1	$f_{ij}^{11} = f_j^1$

(C.9)

f_i^o is the frequency of "0" type answer to the i^{th} item;

f_{ij}^{1o} is the frequency of answers "1" to item i and "0" to item j.

f_{ij}^{11} is the frequency of "1" answers to both items.

The F_{ij} cross tabulation now reads:

(C.10)

		item j: 0	item j: 1	
item i	0	f_i^o	0	f_i^o
	1	f_{ij}^{1o}	f_j^1	f_i^1
		f_j^o	f_j^1	1

From (C.10) we have

$$f_i^o = f_j^o - f_{ij}^{1o} \quad \text{for all } j>i \quad \text{and} \tag{C.11}$$

$$f_i^1 = f_j^1 + f_{ij}^{1o} \quad \text{for all } j>i. \tag{C.12}$$

Recalling that $f_i^o\, w_{io} + f_i^1\, w_{i1} = 0$, we have

$$w_{i1} = -(f_i^o \,/\, f_i^1)\, w_{io}. \tag{C.13}$$

Using (C.11), (C.12) and (C.13) the correlation coefficients are computed as follows:

$$\underline{w}_i' \, C_{ij} \, \underline{w}_j = w_{io}\, w_{jo} \quad f_i^o \,(1 - f_i^o \,/\, f_i^1) = r_{ij} \tag{C.14}$$

$$\underline{w}_i' \, C_{i\ j+1} \, \underline{w}_{j+1} = w_{io}\, w_{j+1\ o} \quad f_i^o \,(1 - f_i^o \,/\, f_i^1) = r_{i\ j+1} \tag{C.15}$$

Hence $r_{i\ j+1} = r_{ij}\,(w_{j+1\ o} \,/\, w_{jo})$, or equivalently $r_{ij} = q\, r_{i\ j+1}$, which is the proposition (C.5) to demonstrate.

The model in which the arrows are inverted reads

$$\begin{aligned} X_3 &= b_{43} X_4 &&+ u_3 \\ X_2 &= b_{42} X_4 + b_{32} X_3 &&+ u_2 \\ X_1 &= b_{41} X_4 + b_{32} X_2 + b_{21} X_2 &&+ u_1 \ . \end{aligned} \tag{C.17}$$

Following a demonstration similar to the above argument, one shows that $b_{ij} = 0$ if $j < i-1$.

ACKNOWLEDGEMENT

This paper belongs under the research supported by the Stiftung Volkswagenwerk.

REFERENCES

Apel, H. and Wold, H., 1982, Soft Modeling with Latent Variables in two or more Dimensions, in Jöreskog K.J. and Wold: Systems under Indirect Observation.

Anderson, T. W., 1958, An Introduction to Multivariate Statistical Analysis (John Wiley and Sons, New York).

Corwin, Lawrence J. and Szczarba, Robert H., 1979, Calculus in Vector Spaces, (Marcel Dekker, New York and Basel)

Efron, B. and Stein, C., 1981, The Jackknife Estimate of Variance, The Annals of Statistics, 3, 586-596.

Goodman, Leo A., 1973, Causal Analysis of Data from Panel Studies and Other Kinds of Surveys, The American Journal of Sociology, 5 ,*reprinted in 1978*

Goodman, Leo A., 1978, Analyzing Qualitative/Categorical Data, (Addison Wesley , London and Amsterdam)

Henning, H. Jörg, Rudinger, Georg, 1983, Hand-out: Workshop "Qualitative Models", 8-12.3.1983, Max-Planck Institut für Bildungsforschung, Berlin.

Horst, Paul, 1965, Factor Analysis of Data Matrices, (Holt, Rinehart and Winston, New York)

Jöreskog, K.J. and Wold, H., 1982, Systems under Indirect Observation. Causality*Structure*Prediction, I-II, (North Holland, Amsterdam)

Kettenring, J. R., 1969, Canonical Analysis of Several Sets of Variables, Institute of Statistics Mimeo Series no 615, University of North Carolina, Chapel Hill, North Carolina.

Lohmöller, Jan-Bernd, 1981, LVPLS 1.6 Program Manual. Latent Variables Path Analysis with Partial Least Squares Estimation, Forschungbericht 81.04, Fachbereich Padagogik, Hochschule der Bundeswehr, München.

Miller, Ruppert G., 1974, The Jackknife - a Review, Biometrika, 61, 1-15.

O'Neill, M. E., 1978, Distributional Expansions for Canonical Correlations from Contingency Tables, Journal of the Royal Statistical Society, 3, 303-312.

Press, S. James, 1972, Applied Multivariate Analysis, (Holt, Rinehart and Winston, New York).

Rao, C. R. and Kamar, M. S. , 1971, Generalized Inverse and its Application, (John Wiley and Sons, New York).

Stone, M., 1974, Cross Validation and Assessment of Statistical Predictions, Journal of the Royal Statistical Society, 2, 111-147.

Upton, J. G. , 1978, The Analysis of Cross Tabulated Data, (John Wiley and Sons, New York).

Wold, H. and Bertholet, J.L., 1983, The PLS Approach to Multidimensional Contingency Tables in A. Rizzi (ed): Transaction of International Meeting on Multivariate Contingency Tables, University of Rome, June 25-26, Metron.

Wold, H. and Bertholet, J.-L., 1982, From Quantitative to Qualitative Data Analysis By PLS (Partial Least Squares) Research Report, Department of Econometrics, University of Geneva Switzerlands

Technical note

Our computer program (FORTRAN simple precision) is the LVPLS program by J.-B. Lohmöller (1981).