JOURNAL OF CHEMOMETRICS, VOL. 2, 231–246 (1988)

NOTES ON THE HISTORY AND NATURE OF PARTIAL LEAST SQUARES (PLS) MODELLING

PAUL GELADI

Norwegian Computing Center, P.B. 114 Blindern, N-0314 Oslo 3, Norway

AND

Research Group for Chemometrics, Umeå University, S-90187 Umeå, Sweden

SUMMARY

The growing awareness of the need for non-deterministic and distribution-free (soft) models combined with an iterative algorithm for finding latent variables led to the construction of partial least squares models. There have been separate developments in the humanistic and the natural sciences, with stress on different aspects and a different terminology. The historical development is described and some key topics are explained.

KEY WORDS Partial least squares models History of PLS Soft versus hard modelling Latent variables Path models

INTRODUCTION

Partial least squares (PLS) is the name for a class of methods, used for relating blocks of variables measured on sets of objects. Modelling with PLS has a large potential as a method of data analysis in many branches of science. This total potential is still rather unused, but PLS is starting to enjoy widespread acceptance among scientists. In this process overview is lost. Widely different branches of science are using identical PLS algorithms with differences in nomenclature and totally different algorithms with varying goals are put in the same basket because of the name PLS.

The pioneering work of PLS was largely done by Professor Herman Wold who worked at different universities in Sweden. This is described in the section on prehistory, spanning the period from the 1930s to the mid 1960s. The section on history describes the developments from the early 1970s to the early 1980s, when the basis for PLS modelling was elaborated and interest grew among econometricians and sociologists. The section on PLS in chemistry describes the introduction and use of one subset of PLS models, mainly comprising regression modelling from 1980 onwards.

The ideas of PLS modelling are closely connected with the evolution of chemometrics as a discipline. Just a small sample from the recent literature, the new journals *Journal of Chemometrics* and *Chemometrics and Intelligent Laboratory Systems*, will show that interest in PLS is very much alive.[1-8]

It is hoped that looking back at the historical development of PLS will bring understanding and transparency to some of the confused users of it. A historical overview can never be complete, but at this stage anything pointing in the right direction may be felt as useful. Because of the historical nature of this text, reference is made to key publications and all the

0886–9383/88/040231–16$08.00

Received 26 October 1987
Revised 11 January 1988

techniques mentioned are not explained completely. Recent literature (after about 1985) has been deliberately left incomplete. One should also be aware of the fact that papers often appear years after the work was carried out.

Besides historical accounts, there are also included sections on basic topics that form the heart of partial least squares reasoning: soft modelling, latent variables and path modelling. A special section is devoted to explaining the differences between PLS philosophies in natural and humanistic sciences.

PREHISTORY

The coming about of partial least squares modelling was a long process, in which different aspects of the method were extracted from other methods (often as a reaction to a method's insufficiency) and refined over a long period of trials (and certainly also of errors).

It all started in the years 1932–1938, when Herman Wold was working on his Ph.D.[9] in the field of time series at the University of Stockholm. Herman Wold's time series work led to a theorem for separating deterministic and stochastic contributions to the time series. He found out that in many examples the deterministic term was absent. This may have been a factor in the choice of non-deterministic models later on.

After receiving his Ph.D., Herman Wold went on to study regression in econometric models. This led to the fixed-point method in the early 1960s. This is a method of designing path models with directly observed variables.[10,11] The algorithms for fixed-point calculations are iterative. This means that an infinite number of iterations is required to give convergence to the desired results. Numerically good results (within the rounding error of the computer) are achieved very quickly, with only a limited number of iterations. This experience with iterative models has played a great role in later developments.

During these early times, Herman Wold gained experience with multivariate modelling and non-deterministic (soft) modelling. Soft models are able to deal with many variables easily, while deterministic models, by their very nature, are limited to a few variables. Figure 1 gives an example of a deterministic (hard) versus a non-deterministic (soft) model. The terms hard and soft are used rather loosely. The ideas of soft and hard in models have their own section further on in the text.

A turning point was around 1964, when Herman Wold invented the NIPALS (non-linear iterative partial least squares) method published in 1966.[12,13] Some of the NIPALS uses (principal components) were a reinvention of the same method that was suggested by Fisher in 1923. The method was first called NILES (non-linear estimation by iterative least square

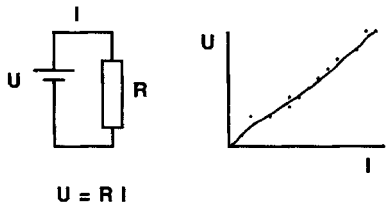

Figure 1. Deterministic and non-deterministic models. Ohm's law gives an exact relationship between the voltage over a resistor and the current flowing through it: a deterministic model. In a real situation there are measurement errors, but even after these are removed a straight line may not be the model of the resistor's behaviour. A polynomial, a Fourier series or a spline may be used as a softer model for electrical resistance. The same also goes for the ideal gas law ($PV=nRT$) and for Lambert–Beer's law ($E=KCl$)

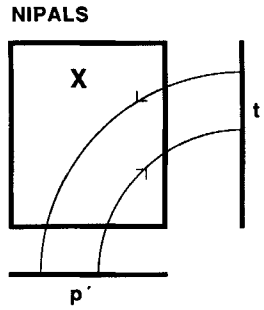

Figure 2. The NIPALS algorithm. A first estimator of the score **t** is multiplied through the data matrix **X** to get an approximation of the loading **p'**. This loading 'proxy' is multiplied back through the data matrix to get a rotated score **t**. After a sufficient number of iterations there is almost always convergence. The results are **t**, the largest eigenvector of **XX'** and its direction in multivariate space, the loading **p'**. This is zero-arrow PLS

procedures). The name NIPALS was used in References 14–18. Herman Wold shows[12] that NILES solves the following problems:

(1) principal components
(2) canonical correlations
(3) hybrid models of principal components, canonical correlation and multiple regression
(4) principal components in the case of partial information
(5) interdependent systems
(6) parameters in the model and in the residual structure
(7) quotient regression
(8) factor analysis
(9) a servo mechanism that involves errors in both the equations and the variables

This list must be seen as a quote and readers are referred to Reference 12 for more details.

The NIPALS method contains a lot of interesting properties, properties that eased the path to useful PLS modelling. NIPALS has also been called: '*how to compute principal components by an iterative sequence of simple ordinary least squares regressions*'. Figure 2 gives a graphical representation of the NIPALS–NILES calculation. Herman Wold mentions[12] that the power method is a special case of NILES. From the same paper comes the following quote about the relation of NILES–NIPALS to maximum likelihood (ML): '*ML requires a full specification of the model, including the stochastic properties of residual deviations. The full specification, making for accurate and refined estimates, is often a premium when the models are linear and the disturbances independent. As is well known ML estimation is liable to get stuck in formidable difficulties when it comes to non-linearities and interdependencies*'.

More details on the early work of Herman Wold can be found in Reference 19.

HISTORY

The combination of econometric modelling and NIPALS led eventually to a first form of PLS (partial least squares) in the early 1970s. It took five years of experimenting before PLS took its definitive shape. Herman Wold gives the end of 1977[20] as the birthdate of PLS.

PLS was first introduced mainly as a path modelling device. It produced latent variables that contained the essentials of the original data and could be used in a simplified path model[15–18].

A peculiarity is that initially PLS models were only used with one latent variable dimension per block. References 17 and 18 give a good overview of how Herman Wold thought of combining NIPALS with some well known path models.

An interesting quote about soft versus hard modelling comes from Reference 15: '*High-information vs. low-information modelling. This is not a dualism, but rather a wide range of diversifications, with many intermediate cases, and with several partings of the ways at fundamental level. Speaking generally, a model gains in operative use and efficiency if prior information is available and is incorporated in the model. On the other hand the model can be overambitious in using information that is not available or is mainly hypothetical*'.

A quote from Reference 18 gives a good idea about the nature of PLS models: '*The model is designed in terms of blocks of manifest (directly observed) variables; each block is approximately represented by a latent (indirectly observed) variable; the path of inner relations between the latent variables are the causal–predictive core of the model; the path may take the form of a causal chain system or an interdependent system; each latent variable is a weighted aggregate of its block of observables, with weights given by the outer relation(s) for this variable; ...*'. A manifest variable is a directly measurable one. A section on latent variables can be found later in the text. A graphical representation is given in Figure 3.

From 1978 on, different types of PLS models found their way in different branches of science.[21,22]

A very important event that probably marks the transition from prehistory to history was the Cartigny Workshop near Geneva, 18–20 October 1979. Its results were presented in two proceedings volumes.[23] The workshop was so successful in generating ideas that it will probably take 20 years before all of them are tried and thoroughly tested. The workshop focused mainly on two types of methods: the distribution-based maximum likelihood approach LISREL (linear structural relation model, Jöreskog–Volume I) and the distribution-free least squares approach PLS (Wold–Volume II). The comparison of distribution-based and distribution-free (PLS) latent variable methods is also highlighted in the articles by Knepel,[24]

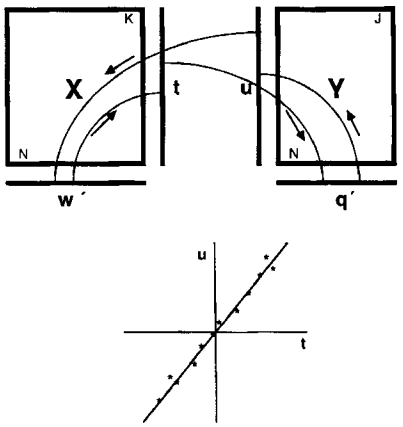

Figure 3. Partial least squares (PLS). A block **X** of K manifest variables measured on N objects is represented by its latent variable **t**. A block **Y** of J manifest variables measured on the same N objects is represented by its latent variable **u**. The relation between the blocks is given by the inner relation between **t** and **u**, the latent variable representations. The latent variables are formed as weighted aggregates of their respective blocks. The weights **w**' and **q**' are obtained from outer relations in their blocks. This is one-arrow PLS. The curved lines and arrows give an idea of how the algorithm works

Fornell and Bookstein[25] and Dijkstra.[26] Lohmöller[27] describes the program LVPLS (latent variables by PLS) and gives some practical information about the differences between PLS and LISREL. He mentions computer resources, non-existing optimality conditions of maximum likelihood for small samples, the non-parametric properties of small finite populations, the possibility of using many variables, exploration without a 'true' model and prediction/prognosis possibilities as assets for PLS.

The proceedings[23] give a lot of examples and algorithms, theory and proofs, but the most interesting parts are the editorials, introductions and conclusions describing the basic philosophy and mental atmosphere.

H. Wold gives a list of branches of science using PLS. He mentions

(1) organic chemistry
(2) analytical chemistry
(3) medical chemistry
(4) education, psychology
(5) management science
(6) economics
(7) political science
(8) environmental science

and gives literature references, most of them in the same proceedings volume.

One of the main properties of PLS that attracted regression users was the fact that the ratio of the number of objects to the number of variables was not constrained. This opened the path to data analysis for a lot of data sets that were previously put aside because they had too many variables and too few objects.

An important aspect of PLS models is that they avoid the trap of collinearity that has always lured users of regression. Key papers are by Jagpal in 1982,[28] S. Wold *et al.* in 1984[29] and Næs and Martens in 1985.[30] PLS as a regression method performs a similar role to principal component regression (PCR), ridge regression (RR) and stepwise multiple linear regression (SMLR).[31]

Another important early aspect of PLS models was the speed of the iterative algorithm. This made PLS rather unique for use on the older minicomputers and some simpler microcomputers. Nowadays, when computing power is going up and prices are going down, the speed aspect may not be so relevant anymore. References 1 and 6 are examples of attempts to describe PLS in a more analytical way.

The use of PLS for analysing higher-order data arrays was mentioned briefly in Reference 13 and in the Cartigny proceedings.[23] A paper describing more details for this methodology by Lohmöller and H. Wold appeared in 1980.[32]

A number of Ph.D. theses having PLS as the main subject have appeared.[33–36]

The earliest known thesis on PLS is that of Hui.[33] He describes the PLS path model, discusses causality and gives examples in education. There are also sections on the comparison of PLS and fixed-point methods and on PLS models for interdependent inner relations. An appendix in this thesis describes computer programs.

Knepel[34] in his thesis mentions factor analysis, principal component analysis, path analysis, LISREL and PLS as methods for econometric modelling. The examples given are from labour market studies.

The thesis of Dijkstra[35] is more theoretical. Dijkstra describes LISREL, introduces PLS and deduces some of its properties, among them convergence.

Table 1. Different types of PLS from the NIPALS algorithm

Name	Number of blocks	Uses
Zero-arrow or zero-order	1	Principal components
One-arrow or first-order	2	Regression
Many-arrow or higher-order	Many	Path modelling

The thesis of Lohmöller[36] is at this moment only available in manuscript form, but a printed version will appear soon.

Work on PLS and multidimensional contingency tables was carried out by H. Wold and Bertholet.[37,38] Studies on the use of PLS for path modelling in education studies were done by Noonan and H. Wold.[39–41]

There has been confusion as to which models deserve the name PLS. A useful classification would be to count the number of arrows. There seems to be a very profound philosophical reason for choosing a certain model with a certain number of arrows. Using the NIPALS algorithm for calculating principal component latent variables then becomes 'zero-arrow PLS', two-block regression becomes 'one-arrow PLS' and more complex path models become 'many-arrows PLS'. Other names might be 'zero-order PLS', 'first-order PLS' and 'higher-order PLS'. An overview is given in Table 1. It is important to realize that there is a quantum jump in philosophy between the orders of PLS.

PLS IN CHEMISTRY

PLS found its way into chemistry in the late 1970s. The first fields to use it were organic and analytical chemistry. These fields are characterized by their production of multivariate data and their lack of (working) physical models. The fields of physical and inorganic chemistry, where chemistry is often identical with the study of physical relationships for a few variables, did not follow. Because of this introduction in chemistry, a lot of related fields were also affected: biology, clinical chemistry, medicine, food research, biotechnology, quantitative–structure activity relationships (QSAR), pharmacology, etc.

An early paper mentioning the possibilities of relating many blocks of data by latent variables was by Svante Wold, the son of Herman, in 1977.[42] Svante Wold had helped his father in the previous work on the NIPALS algorithm and used this in his own work on the SIMCA classification method.[43–45] The following quote comes from Reference 42: '*By means of the so called soft-C models, each of the blocks of variables is approximated by a separate PC model but, in addition, linear relations are introduced between latent variables in the different blocks ...*'.

The first chemical paper to make reference to PLS was by Gerlach, Kowalski and H. Wold in 1979,[46] given in a more elaborate version in Reference 23. It was a paper on path modelling of river pollution. The main interest for PLS in chemical applications was regression between two blocks. An early explanation of the philosophy is given by Martens and Jensen[47] and S. Wold, Martens and H. Wold.[48] An important reason for using PLS is the replacement of univariate wet chemistry methods (needing only hypothesis testing and linear regression) by multivariate instrumental methods (chromatography, spectrometry).

Already from the start there was a difference between two schools: that of Svante Wold and that of Harld Martens. The Svante Wold algorithm[48] for regression is based on orthogonal

latent variables, which is a numerical advantage. The Harald Martens algorithm[47] uses slightly correlated latent variables, the advantage being that an easier explantion of statistical properties is possible. Both algorithms have exactly the same predictive properties.

In 1983 a lot of papers of PLS in analytical and organic chemistry appeared.[49–52] Around the same time, Magni Martens and co-workers started using PLS for relating sensory variables to chemical variables in food science.[53,54]

In the autumn of 1982 (20–23 September) there was a meeting in Oslo called 'Food Research and Data Analysis'. The meeting was not primarily about PLS, but many of the early PLS users in chemistry were present and PLS is often mentioned in the proceedings.[55]

Something similar to the Cartigny Workshop mentioned above happened from 4 to 8 June 1984 in Umeå. At a meeting called 'MULDAST' (multivariate data analysis in science and technology) specialists in the fields of data analysis and chemical applications gathered to discuss PLS and related problems. There are no extended proceedings available from this meeting.[56]

As for the use of regression, one could state that there are three basic models:

(1) causal regression with a known but imperfect underlying model (as in Beer's law modelling in spectrometry)
(2) inverse causal regression (as in Beer's law prediction) where the causal pathway is reversed for practical reasons
(3) exploratory regression where there is no know causal pathway or underlying theory (as in relating sensory responses to chemical and physical properties of food products)

The PLS regression model that has gained high popularity in chemical and related applications can be written as follows:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{f} \tag{1}$$

or

$$\mathbf{Y} = \mathbf{XB} + \mathbf{F} \tag{2}$$

Equation (1) gets the name 'PLS1' and equation (2) is called 'PLS2'.

$\mathbf{y}$ a vector or responses ($N \times 1$)
$\mathbf{Y}$ a matrix of responses ($N \times J$)
$\mathbf{X}$ a matrix of measured features ($N \times K$)
$\mathbf{b}$ a vector of PLS regression coefficients ($K \times 1$)
$\mathbf{B}$ a matrix of PLS regression coefficients ($K \times J$)
$\mathbf{f}$ a vector of residuals($N \times 1$)
$\mathbf{F}$ a matrix of residuals ($N \times J$)
N number of objects, samples
K number of predictor variables
J number of responses

Here $\mathbf{X}$ and $\mathbf{Y}$ are the same as those shown in Figure 3. Reference is made to the usual regression literature.[31] The difference in PLS is in the way $\mathbf{b}$ (equation (1)) or $\mathbf{B}$ (equation (2)) are calculated from latent variables.[47,48]

The causal regression types are good for the calibration of instruments in laboratory and industry. This is known as multivariate calibration. The Ph.D. thesis of Martens[57] is a good source of information for this kind of activity. Many articles of Martens and co-authors are

reprinted in this thesis and are therefore not mentioned separately in the reference list in the present paper.

Usually, causality is not considered in the setting up of PLS regression models, but the variable(s) to be predicted are put in the dependent block. If no prediction is desired, the block with the lowest expected rank is postulated dependent for practical reasons.

Exploratory use of PLS can be made when no physical models are known that link independent and dependent variables. This is often the case in the biological, medical and related fields.

Recently users of PLS became aware of the possibility of designing the blocks of variables instead of just filling them with haphazardly collected data. An early publication on the combination of designed blocks and PLS was by S. Wold, Dunn and Hellberg.[58]

A paper mentioning the need for analysing higher-order data arrays by Esbensen and S. Wold appeared in 1983.[59]

HARD AND SOFT MODELS

There is some confusion about the use of the terms hard and soft, and maybe even on the meaning of the term model. The distinction should be made between the statistical and the physical model.

A physical model is a mathematical relation, a function, with variables and parameters. As an example, Beers law can be considered as

$$A = KC \tag{3}$$

where A is an absorbance, C a concentration and K the extinction coefficient. This hard model is not very often found and many manufacturers build other functions into their instruments. A general expression would be

$$A = A_0 + KC + f(C) \tag{4}$$

where A_0 is a constant term (offset), an absorbance without the absorbing chemical present, and $f(C)$ is a function describing non-linear behaviour. There are many ways of writing $f(C)$, and all of them have advantages and disadvantages, situations where they work and situations where they do not work. As a physical model of a phenomenon, equation (3) can be called hard and equation (4) soft. A lot of the early PLS work consisted of discovering this difference for econometric models. There is a whole range in models between real hard, physical deterministic and soft. As usual, the extremes in the range are almost never useful.

A new complication arises when error in measurement is considered. In this case A in equations (3) and (4) is a stochastic variable. This can also be written

$$A = A_0 + KC + f(C) + e \tag{5}$$

where e is an error term. This equation can be called a statistical model; e is considered as belonging to a distribution. A well known one is the normal distribution

$$e \in N(0, \sigma^2) \tag{6}$$

When the distribution of e and its parameters are well known, a whole battery of statistical hypothesis tests can be carried out. This is a hard statistical model. But sometimes this distribution is unknown or its parameters cannot be estimated, or one of many other complications exists. Working with only partly known distributions or distribution parameters without ignoring the existence of e can be called a soft statistical model. This forms the main

Table 2. Soft and hard aspects of models

Aspect of model	Hard model	Soft model
Physical	Deterministic 'law'	Empirical
Statistical	Distribution model	Distribution-free
Causality	Strictly causal	Correlation
Number of variables	Univariate	Multivariate

difference between the maximum-likelihood-based LISREL and the least-squares-based PLS philosophy. Here also the extremes in the range between hard and soft are rare occurences.

Other aspects of the range between hard and soft can be mentioned. The acceptance of a causal model can be called hard, while studying correlation structure when causality is unknown is a softer variant. In chemistry, biochemistry and medicine, QSAR (quantitative structure–activity relationships) is a typical example where causal models are unknown. Then there is also the step from univariate to multivariate that can be called a 'softening' of the model. This means that, even if there is a univariate base model, corrections from many other variables must be taken into account.

All these aspects are collected in Table 2. In many cases there is a mixture of aspects and the extremes of purely hard or purely soft model are never reached.

LATENT AND MANIFEST VARIABLES

The concepts of manifest and latent variables are at the basis of multivariate thinking and therefore also at the basis of PLS thinking. The concept of a latent variable is a very natural one. Latent variables are intuitive, hidden, summarizing variables and not directly measurable. Manifest variables are directly measurable, but may have no meaning for the phenomenon or problem under study. In univariate situations the distinction between latent and manifest variables makes no sense. It only appears in multivariate situations. Some examples from the classical PLS literature will be used to explain the difference. The vast literature on factor analysis, principal component analysis and related topics is concerned with latent variables. A good overview article on latent variables in LISREL and PLS was written by Bookstein.[60]

References 21 and 28 give an example from advertising as a marketing device. The manifest variables are monthly radio advertising and monthly newspaper advertising. They are augmented by non-linear transformations. The unobservable latent variable is awareness. Awareness leads to a new latent variable, preference. Preference cannot be observed, but monthly sales of the advertized products can. This leads to the PLS model shown in Figure 4.

A chemical example comes from Martens.[57] Near infrared spectra of food products (meat) are measured. The manifest variables are the absorbances at different wavelengths. The desired variable may be the protein content of a sample. None of the manifest variables is really good at giving information about the protein content. With the help of PLS, it is possible to construct a latent variable that is a very good predictor of protein content.

The latent variable can be used in different ways. In the first example, studying awareness and preference and their relation may be the ultimate goal of the study. Another goal may be the prediction of sales. In this case the latent variables are just a tool for achieving a predictive model. Yet another goal may be to assess whether one of the manifest variables is unimportant

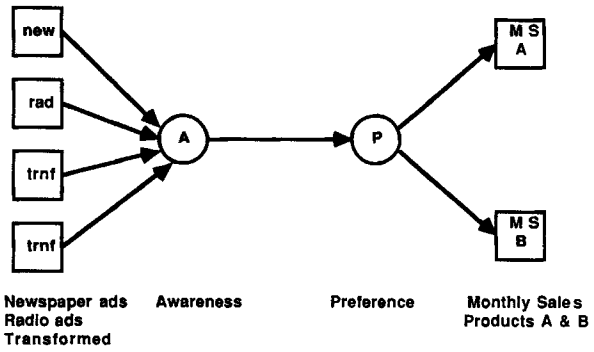

Figure 4. The arrow scheme for a PLS model. Squares indicate manifest variables (NEWspaper ads, RADio ads, TRaNsFormed variables; Monthly Sales, A, Monthly Sales B), circles indicate latent variables (Awareness, Preference). The arrows between the latent variable circles symbolize the inner relation. The scheme shown here is the simplest one possible. Configurations with three and more latent variables have been proposed. The direction of the arrows is an expression of assumed causality

in constructing the latent variable and can be left out (in this study it turned out that expensive radio advertisements could be dropped).

For the calculation of latent variables, Herman Wold proposed two basic methods: mode A and mode B. They can be described as follows:

$$\mathbf{t} = \mathbf{X}\mathbf{w} \qquad \text{mode A} \tag{7}$$

$$\mathbf{t} = \mathbf{X}\mathbf{b} + \mathbf{e} \qquad \text{mode B} \tag{8}$$

In mode A the latent variable **t** is found by projecting the block of manifest variables **X** onto a direction **w** in space. In mode B the latent variable **t** and the block **X** of manifest variables are connected by a multiple linear regression with regression coefficients **b**. It should be emphasized that equations (7) and (8) only make sense in an iterative algorithim. Since the choice of mode A or mode B for every block can be different, mixed modes are possible where one block of manifest variables gets mode A and the other one mode B. This is the mixed mode called C. In the natural sciences mode A algorithms have been used almost exclusively.

The meaning of the latent variables can also depend on the type of PLS used. For zero-arrow PLS the latent variables are purely summarizing. For one-arrow PLS the latent variables are partly summarizing and partly correlation-maximizing (here the correlation between the latent variables for blocks is meant). This is in contrast to PCR where the latent variables are only summarizing. For many-arrow PLS the tradition has been to let the role of the latent variable be decided by some *a priori* knowledge of the system under analysis.

It should be stressed here that latent variables are estimated as *linear* combinations of the manifest variables. In many cases it turns out that not one latent variable but a subspace spanned by a number of latent variables (one could call this the latent subspace) explains a phenomenon. In this case it may be possible to find a linearizing transformation that reduces this subspace is just one latent variable. If the linearizing transformation is not known in advance, the space spanned by the manifest variables can be augmented with well known linearizing transformations. The best transformations will then automatically get a high weight in forming the latent variable.

In the first example the manifest variables were

x_1 monthly radio advertising

x_2 monthly newspaper advertising

The transformed variables formed from these were

$$\ln(x_1) \quad \ln(x_2) \quad \ln(x_1)\ln(x_2) \quad \ln(x_1)^2 \quad \ln(x_2)^2$$

The latent variable t_1 then takes the shape

$$\ln(t_1) = a_0 + a_1\ln(x_1) + a_2\ln(x_2) + a_{11}\ln(x_1)^2 + a_{22}\ln(x_2)^2 + a_{12}\ln(x_1)\ln(x_2)$$

where $a_{..}$ are weighting constants.

In the second example[57] a transformation based on known spectrometric properties of the samples and the measurement instrument was used to reduce the latent subspace from about a seven-dimensional subspace to one latent variable.

PATH MODELS

Path analysis was first fully described in 1934 by Sewall Wright, then working at the Zoology Department of the University of Chicago.[61] The method was used for studying genetic paths in breeding. The first path models were concerned with causal relations and therefore often called causal path models. Wright[61] gives a discourse on whether causality really exists and on the possible reversability of time.

The original idea of Sewall Wright was to get a consistent system for explaining correlation coefficients between two variables connected by a path when some correlation coefficients along the path were known. This is explained very well by Li.[62] Later uses of path modelling concentrated less on correlation coefficients, but used regression relations between the (blocks of) variables.

Later on feedbacks and bidirectional paths were included in path modelling. This was necessary for explaining sociological and econometric models. These types of models containing bidirectional paths and feedbacks are called interdependent models. The method of path modelling was introduced in the 1960s in sociology and found applications in econometrics in the 1970s. According to Herman Wold,[20] it was Otis Dudley Duncan who in the early 1960s proposed the merging of path analysis and latent variables into one method. In the early days these latent variables were calculated for each block separately. The contribution of PLS is to let the blocks of manifest variables know about each other's latent variables for constructing the latent variables. This optimizes the inner relations.

The topic of path analysis[63–65] is concerned with studying relations between blocks of manifest variables. These relations can be very complex and are represented by an arrow scheme. Because of the complexity of construction and difficulties in understanding and explaining such arrow schemes, it is not surprising that only one latent variable was considered sufficient in many examples.

Path analysis seems to lack its analogy apart from genetics in the natural sciences. It is not certain whether the methods applied in sociology and econometrics can be used directly to solve problems in the natural sciences. One field where path models could be made useful is the process industry, where causal pathways and feedbacks are present. The objects in this case are samples measured at fixed time intervals in the different stages of the process.

It has been shown that, in order to have a regression model with good predictive properties,

it is best to put all predictor variables in one large block and all response variables in another block, without worrying too much about any special paths. Hence it can be concluded that path modelling is more a tool for exploring existing data and that its use in prediction can be limited.

Illustrations of the path models studied by PLS users can be found in many of the literature references. Figure 5 gives an explanation of some of the ideas encountered in path modelling. These are only a fraction of all possible path models. Some very complex path models and a paragraph on path modelling history can be found in Reference 40.

SOME DIFFERENCES BETWEEN PLS MODELS IN NATURAL AND HUMANISTIC SCIENCES

There has always been a confusion on how PLS was used in natural and humanistic sciences. PLS is very versatile and allows different answers to different questions.

Natural sciences often require calibration with subsequent prediction where an underlying physical model is present. This physical model can be something like Beer's law or other spectroscopic laws. These physical laws are mental constructions that are never obeyed strictly and always influenced by more variables than expected. Natural sciences may also have low noise in their measurements and require precise and accurate results.

Humanistic sciences are generally known for their very noisy measurements for a situation where no underlying model is expected in advance. Hence the use is exploratory, often with no demands for future prediction. In this case the extraction of one or two useful latent variables is a great success. This gives a larger freedom to experiment with different PLS-type models. The humanistic scientists have also been very good at combining the basic principles of path modelling with their PLS models.

There also seems to be a better awareness of causality in the humanistic sciences, whereas causality is used in a very sloppy way in natural sciences. This is shown by the importance given to the arrow scheme and the choice of the direction of the arrows in this scheme (Figure 5).

Many applications fall between the two cases mentioned above. The sensory data analysis mentioned above[53,54] is an example of a bridge between precisely measured instrumental data and psychological data.

The high precision required by regression and calibration in the natural sciences has led to special attention to methods for determining the number of factors required. This can be compared in difficulty with the rank estimation problem in PCA.

SUMMARY

PLS methodology is still in development. Many recent publications of high quality and sophistication were not included here because their content and philosophy did not have time enough to make an impact on the scientific community. Important future events may be the use of two-block PLS regression by traditional path model (the Herman Wold school) users and the extension of the two-block PLS regression philosophy (the Svante Wold and Harald Martens school) to many blocks and eventually to path models.

It is almost impossible to summarize what has been going on during the history of PLS. Probably the best way is by giving a list of topics of concern in the PLS papers mentioned. It was

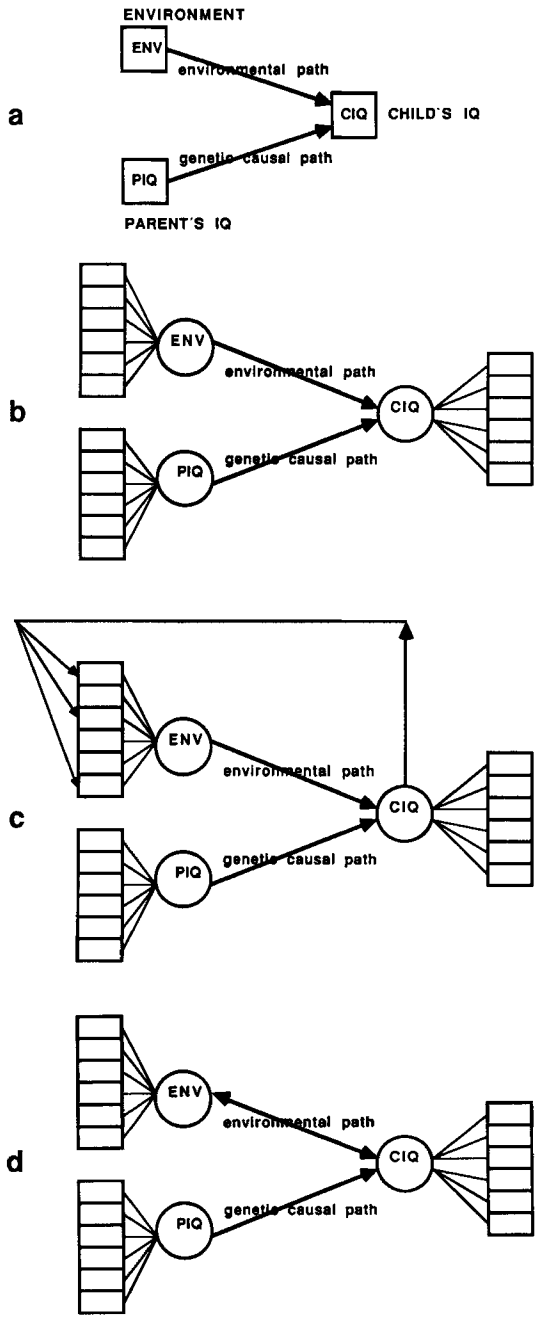

Figure 5. Path models. (a) A simple model described by Sewall Wright.[61] It is a causal model for manifest variables (rectangles). The genetic path from the parent's intelligence to the child's intelligence is certainly causal and irreversible. Also the environmental path can be considered as irreversible for practical reasons. (b) A natural extension of the type introduced by Otis Dudley Duncan. The IQs of both parents and child can be measured by many manifest variables just like the environmental parameters. The latent variables 'IQ of child', 'IQ of parents' and 'environment' are given as circles. This is a simple many-arrow PLS model. (c) An intelligent child might start influencing some aspects of its own environment, thereby causing a feedback loop in the path model. (d) The feedback loop in (c) can be replaced by an interaction (double arrow) between the two latent variables

found appropriate to divide this list into two: one for the humanities and econometrics field and one for the natural sciences field.

For the *humanities* and *econometrics* field:

(1) non-deterministic modelling
(2) many variables
(3) latent variables
(4) iterative algorithms
(5) linear models for non-linear phenomena
(6) path analysis
(7) data-rich and model-poor situations

For the *natural sciences* field:

(1) instrumentation leads to many variables
(2) correlated, collinear variables
(3) designs to span the object space
(4) 'right' number of factors
(5) graphical representation of results
(6) bilinear modelling

ACKNOWLEDGEMENTS

This article is based on a lecture given at the symposium 'Data Approximation and PLS modeling', Oslo, 19 May 1987. Many friends and colleagues sent in interesting material and helped with useful comments and hints. The list has become to long to write it out completely. A special thanks goes to Herman Wold of the University of Uppsala, Sweden. In spite of his advanced age he volunteered enthusiastically to provide a lot of useful information.

REFERENCES

1. A. Lorber, L. Wangen and B. Kowalski, *J. Chemometrics* **1**, 19 (1987).
2. S. Wold, P. Geladi, K. Esbensen and J. Öhman, *J. Chemometrics* **1**, 41 (1987).
3. L. Ståhle and S. Wold, *J. Chemometrics* **1** , 185 (1987).
4. N. Vogt, *Chemometrics Intell. Lab. Syst.* **1**, 213 (1987).
5. I. Frank, *Chemometrics Intell. Lab Syst.* **1**, 223 (1987).
6. R. Manne, *Chemometrics Intell. Lab. Syst.* **2**, 187 (1987).
7. T. Karstang and R. Eastgate, *Chemometrics Intell. Lab. Syst.* **2**, 209 (1987).
8. K. Esbensen, and H. Martens, *Chemometrics Intell. Lab. Syst.* **2**, 221 (1987).
9. H. Wold, *A Study in the Analysis of Stationary Time Series*, Almqvist & Wiksell, Stockholm (1938).
10. H. Wold, *Ark. Mat.* **6**, 209 (1965).
11. H. Wold, *Ark. Mat* **6**, 220 (1965).
12. H. Wold, in *Research Papers in Statistics*, ed. by F. David, p. 411, Wiley, New York (1966).
13. H. Wold, in *Multivariate Analysis*, ed. by P. Krishnaiah, p. 391, Academic Press, New York (1966).
14. H. Wold and E. Lyttkens, *Bull. Int. Statist. Inst.* **43**, 29 (1969).
15. H. Wold, *Eur. Econ. Rev.* **5**, 67 (1974).
16. H. Wold, in *Perspectives in Probability and Statistics: Papers in Honor of M.S. Bartlett*, ed. by J. Gani, p. 117, Academic Press, London (1975).
17. H. Wold, in *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modeling*, ed. by H. Blalock, A. Aganbegian, F. Borodkin, R. Boudon and V. Capecchi, p. 307, Academic Press, New York (1975).
18. H. Wold, in *Quantitative Wirtschaftsforschung: Wilhelm Krelle zum 60. Geburtstag*, ed by H. Albach, E. Helmstedter and R. Henn, p. 729, Mohr, Tübingen (1977).

19. H. Wold, in *The Making of Statisticians*, ed. by J. Gani, p. 190, Springer, New York (1982).
20. H. Wold, in *A Second Generation of Multivariate Analysis*, ed. by C. Fornell p. 325, Preager, New York (1982).
21. B. Hui and H. Jagpal, *American Institute for Decision Sciences: Proc. Ann. Meeting*, Vol 1–2, p. 71 (1979).
22. F. Bookstein, in *Evaluation of Econometric Models*, ed, by J. Kmenta and J. Ramsey p. 75, Academic Press, New York (1980).
23. K. Jöreskog and H. Wold. (eds), *Systems under Indirect Observation: Causality, Structure, Prediction*, North-Holland, Amsterdam (1982).
24. H. Knepel. *Statist. Hefte* **22**, 248 (1980).
25. C. Fornell and F. Bookstein, *J. Marketing Res.* **11**, 440 (1982).
26. T. Dijkstra, *J. Econometrics* **22**, 67 (1983).
27. J. Lohmöller, in *'Statistik– Software', 3e Konf. uber die anwendung von Statistik-Software, Nürnberg*, ed. by W. Lehmacher, p. 179 (1985).
28. H. Jagpal, *J. Marketing Res.* **11**, 431 (1982).
29. S. Wold, A. Ruhe, H. Wold and W. Dunn, *SIAM J. Sci Statist. Comput.* **5**, 735 (1984).
30. T. Næs and H. Martens, *Commun. Statist. – Simul. Comput.* **14**, 545 (1985).
31. N. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York (1981).
32. J. Lohmöller and H. Wold, 'Three-mode path models with latent variables and partial least squares (PLS) parameter estimation', presented at *Eur. Meeting of the Psychometrics Society, The Netherlands* (1980).
33. 'B. Hui, The partial least squares approach to path models of indirectly observed variables with multiple indicators', *Ph.D. Thesis,* University of Pennsylvania, Philadelphia (1978).
34. H. Knepel, 'Sozioökonomische Modellbildung mit unbeobachtbaren Variablen', *Ph.D. Thesis*, Technical College, Darmstadt (1980).
35. K. Dijkstra, 'Latent variables in linear stochastic models: Reflections on maximum likelihood and partial least squares methods', *Ph.D. Thesis*, University of Gröningen (1981).
36. J. Lohmöller, 'Path models with latent variables and partial least squares (PLS) estimation', *Dissertation*, Hochschule der Bundeswehr, Neubiberg; personal communication from the author (1983).
37. H. Wold and J. -L. Bertholet *Metron* **60**, 303 (1982).
38. J. -L. Bertholet and H. Wold, in *Measuring the Unmeasurable*, ed. by P. Nijkamp p. 253, Martinus Nijhoff, The Hague (1985).
39. R. Noonan and H. Wold, *Scand. J. Educ. Res.* **24**, 1 (1980).
40. R. Noonan and H. Wold, *Eva. Educ. Int. Rev. Ser.* **7**, 219 (1983).
41. R. Noonan and H. Wold, in *International Encyclopedia of Education*, ed. by T. Husén and T. Postlewaithe p. 3769, Pergamon Press, Oxford (1984).
42. S. Wold in *Proc. First Int. Symp. on Data Analysis and Informatics, Versailles*, p. 683 (1977).
43. S. Wold, *Pattern Recognition* **8**, 127 (1976).
44. S. Wold and M. Sjöstrom, in *Chemometrics: Theory and Applications*, ed by B. Kowalski, *Amer. Chem. Soc. Symp. Ser. 52*, p. 243 (1977).
45. H. Wold, in *Mathematical Economics and Game Theory: Essays in Honor of Oscar Morgenstern*, ed. by R. Henn and O. Moeschlin, p. 536, Springer, Berlin (1977).
46. R. Gerlach, B. Kowalski and H. Wold, *Anal. Chim. Acta* **112**, 417 (1979).
47. H. Martens and S. -A. Jensen, in *Progress in Cereal Chemistry and Technology*, ed. by J. Holas and J. Kratochvil p. 607, Elsevier, Amsterdam (1983).
48. S. Wold, H. Martens and H. Wold, in *Matrix Pencils*, ed. by A. Ruhe and B. Kågström, p. 286, Springer, Heidelberg (1983).
49. W. Lindberg, J. -Å. Person and S. Wold, *Anal. Chem.* **55**, 643 (1983).
50. M. Sjöstrom, S. Wold, W. Lindberg, J. -Å. Persson and H. Martens, *Anal. Chim. Acta* **150**, 61 (1983).
51. I. Frank, J. Feikema, N. Constantine and B. Kowalski, *J. Chem. Inf. Comput. Sci.* **24**, 20 (1983).
52. I. Frank, J. Kalivas and B. Kowalski, *Anal. Chem.* **55**, 1800 (1983).
53. M. Martens, H. Martens and S. Wold, *J. Sci. Food Agric.* **34**, 715 (1983).
54. M. Martens, B. Fjeldsenden, H. Russwurm and H. Martens, in *Sensory Quality in Foods and Beverages: Definition, Measurement and Control*, ed. by A. Williams and R. Atkin, p. 233, Ellis Horwood Ltd, Chichester (1983).

55. H. Martens and H. Russwurm (eds), *Food Research and Data Analysis*, Applied Science, London (1983).
56. S. Wold, personal communication (1985).
57. H. Martens, 'Multivariate calibration: quantitative interpretation of non-selective chemical data', *Ph.D. Thesis*, University of Trondheim (1985).
58. S. Wold, W. Dunn and S. Hellberg, in *Drug Design, Fact or Phantasy*, ed. by G. Jolles and K. Woolridge, p. 95, Academic Press, London (1984).
59. K. Esbensen and S. Wold, in *Proc. Symp. on Applied. Statistics, Stavanger*, p. 11 (1983).
60. F. Bookstein, in *Advances in Developmental Psychology*, ed. by M. Lamb, A. Brown and B. Rogoff, Vol 4, p. 203, Lawrence Erlbaum Associates, Hillsdale, New Jersey (1986).
61. S. Wright, *Ann. Math. Statist.* **5**, 161 (1934).
62. C. Li, *Population Genetics*, p. 144, University of Chicago Press, Chicago (1955).
63. O. Duncan, *Amer. J. Sociol.* **72**, 1 (1966).
64. H. Costner (ed.), *Sociological Methodology*, Jossey-Bass, San Francisco (1971).
65. H. Blalock, A. Aganbegian, F. Borodkin, R. Boudon and V. Capecchi (eds), *Quantitative Sociology: International Perspectives on Mathematical and Statistical Modelling*, Academic Press, New York (1975).