SYSTEMS ANALYSIS BY PARTIAL LEAST SQUARES (PLS)

Herman Wold

September 1983
CP-83-46

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
A-2361 Laxenburg, Austria

FOREWORD

This Collaborative Paper is one of a series embodying the outcome of a workshop and conference on Economic Structural Change: Analytical Issues, held at IIASA in July and August 1983. The conference and workshop formed part of the continuing IIASA program on Patterns of Economic Structural Change and Industrial Adjustment.

Structural change was interpreted very broadly: the topics covered included the nature and causes of changes in different sectors of the world economy, the relationship between international markets and national economies, and issues of organization and incentives in large economic systems.

There is a general consensus that important economic structural changes are occurring in the world economy. There are, however, several alternative approaches to measuring these changes, to modeling the process, and to devising appropriate responses in terms of policy measures and institutional redesign. Other interesting questions concern the role of the international economic system in transmitting such changes, and the merits of alternative modes of economic organization in responding to structural change. All of these issues were addressed by participants in the workshop and conference, and will be the focus of the continuation of the research program's work.

Geoffrey Heal
Anatoli Smyshlyaev
Ernő Zalai

SYSTEMS ANALYSIS BY PARTIAL LEAST SQUARES (PLS)

Herman Wold*

1. Introduction

The advent of the computer in the early 1950's marks a new era in Systems Analysis.

The first wave: Trend simulation, by the Club of Rome and other teams, was launched with high aspirations and expectations. The performance left much to be desired, and in 1981 a well-documented appraisal stated that the high expectations had not materialized; Kappel & Schwarz, 1981.

In the meanwhile two lines of systems analysis had come to the fore:

Path models with manifest (directly observed) variables, and

Path models with latent (indirectly observed) variables.

Fg.1 shows arrow schemes for two arrays of path models, to the left models with MVs (manifest variables), to the right models with LVs (latent variables). The models II-V to the left are classical in Econometrics. The models I* - II* to the right have their origin in Psychometrics. Models III* - V*, mergers of Econometrics and Psychometrics, were introduced in Sociology in the mid-1960's.

* Uppsala University and University of Geneva

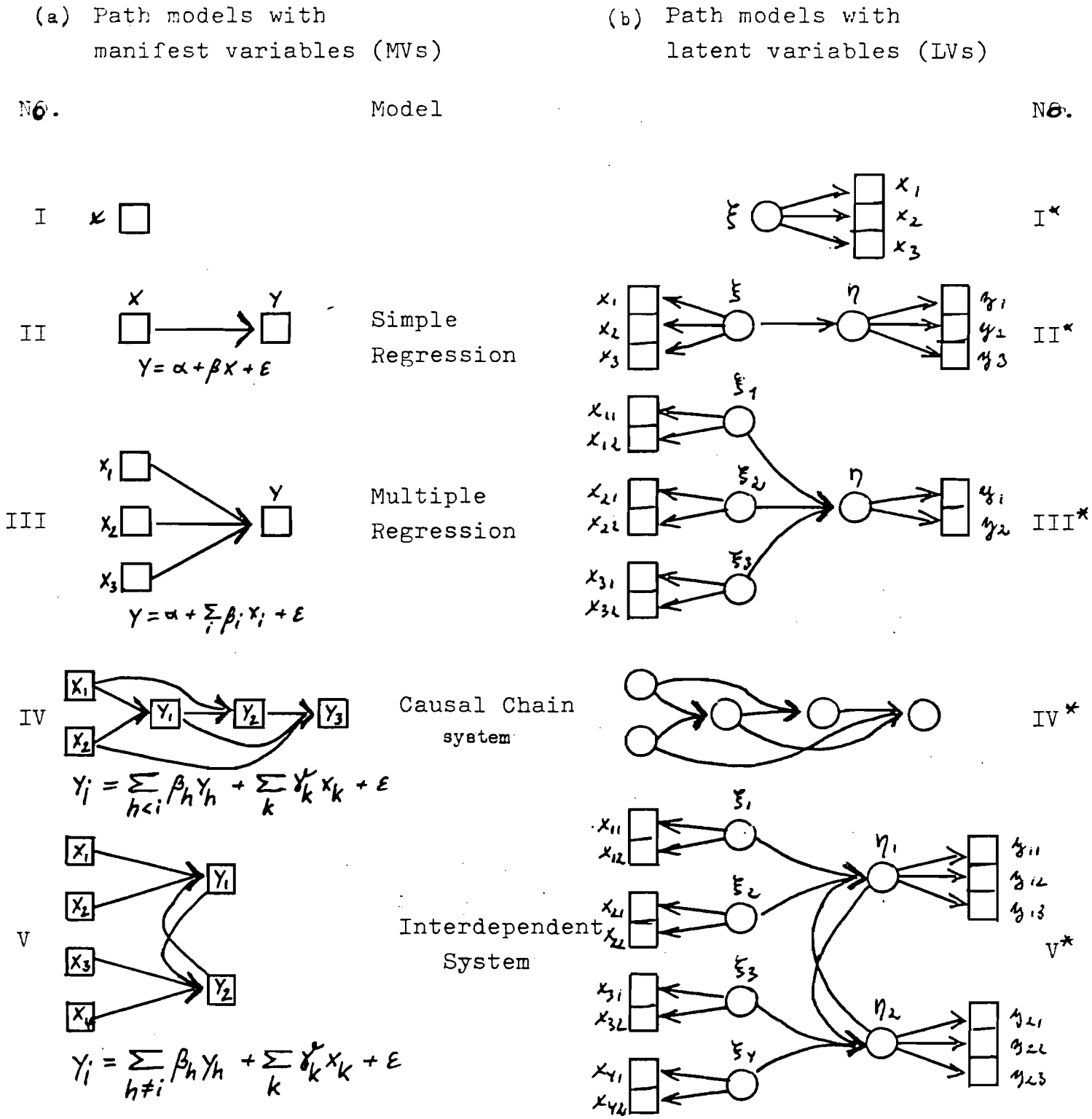

Figure 1 a,b — Graphic illustrations of path models with (a) directly observed variables, and (b) latent variables indirectly observed by multiple indicators.

Sn.2 considers the models to the left in Fg.1, path models with MVs.

Models with one relation (the path has one step):

II. Simple OLS (Ordinary Least Squares) regression;

III. Multiple OLS regression.

Models with two or more relations (the path has more than one step):

IV. Causal Chain (also called Recursive) systems;

V. Interdependent (ID) systems.

Models IV-V are well-known from macroeconomic model building. With reference to the subsequent formulas (1a)-(1b), Causal Chains and ID systems are defined by their structural form (SF), and their reduced form (RF) gives the endogenous variables in terms of the exogenous variables. OLS regression is consistent when applied to Causal Chain systems, but not in the estimation of ID systems.

Sn.3 sets forth the FP approach to ID systems (V). Sn.4 reviews the principles of model building, with focus on ML (Maximum Likelihood) versus LS (Least Squares) modeling. Sn.5 outlines the evolution of the psychometric Models I^* - II^* in Fg.1. Sn.6 reviews the recent advent of general estimation of Models III^* - V^*, namely K.G. Jöreskog's ML algorithm LISREL, followed by H. Wold's LS algorithm PLS.

Sn.7 shows the basic design of PLS modeling, Sn.8 some of its generalizations; Sn. 9 is a brief discourse on applied with with PLS, and Sn. 10 gives a concluding outlook.

2. Path models with manifest variables (MVs).

Jan Tinbergen in his pioneering work with Causal Chain systems, 1935-39, estimated the SF relations by OLS regression. Trygve Haavelmo in 1943 introduced Simultaneous Equations (1 a-b); claiming that OLS is inconsistent when applied to their SF, he recommended estimation by ML (Maximum Likelihood) estimation. Bentzel Wold (1946) distinguished between Recursive systems (IV) and Nonrecursive systems (V), and showed that ML and OLS give numerically the same parameter estimates when applied to the SF of Recursive systems IV. To improve the general rationale of LS (Least Squares) estimation, H. Wold (1958-63) introduced the notion of predictor specification, assuming that the systematic part of the relation to be estimated is the conditional expectation of the target variable; cf. (1c) and (2 a-b).

2.1 Formal aspects. Causal Chains IV and ID systems V have exogenous variables $x = (x_1=1, x_2, \ldots, x_m)$ and endogenous variables $y = (y_1, \ldots, y_n)$. The models are defined by their structural form (SF) which, when solving for the endogenous variables, gives the reduced form (RF).

$$\text{SF:} \qquad y = \beta y + \Gamma x + \zeta \qquad (1a)$$

$$\text{RF:} \qquad y = [I - \beta]^{-1}\Gamma x + \varepsilon \qquad (1b)$$

Matrix notation: $\beta = [\beta_{ik}]$; $\Gamma = [\gamma_{ih}]$, $\qquad i, k = 1, n;\ h = 1, m$

2.2 Comparative aspects of Models IV and V:

IV. Causal Chain (Recursive) systems — V. Interdependent (ID) systems

Matrix β is triangular, $\beta_{ik} = 0$, $k \geq i$. — β is nontriangular, $\beta_{ik} \neq 0$ for some $k > i$.

Joint feature: $E(y_i \mid x_1, x_2, \ldots, x_n) = [I - \beta]^{-1} \Gamma x$ (1c)

" " — OLS estimation of RF is consistent

$E(y \mid y, x) = \beta y + \Gamma x$ (2a) — $E(y \mid y, x) \neq \beta y + \Gamma x$ (2b)

OLS estimation of SF is consistent — OLS estimation of SF is inconsistent.

For consistent estimation of the SF in ID systems the TSLS (Two Stage Least Squares) method was introduced by H.Theil (1953, 1958) and R.Basmann (1957), on the Classical ID assumption that each residual ε (or ϑ) is uncorrelated with all exogenous variables x.

3. The Fix-Point approach. In the Fix-Point (FP) method for the estimation of ID systems (H. Wold, 1965-66) the key feature is to reformulate the SF by replacing each explanatory endogenous variable y by its conditional expectation, denoted by y^*, first adopting and then generalizing the Classical ID assumption.

REID (REformulated ID) systems: each residual ε (or ϑ) is assumed to be uncorrelated with all exogenous variables x;

GEID (GEneral ID) systems: in the i:th relation of the SF ($i = 1, n$) the residual ε_i is assumed to be uncorrelated with all variables y^*, x that occur in the i:th relation.

In symbols, REID and GEID systems read as follows.

SF: $y = \beta y^* + \Gamma x + \varepsilon$ (3a)

RF: $y = [I - \beta]^{-1} \Gamma x + \varepsilon$ (3b)

$y^* = [I - \beta]^{-1} \Gamma x$ (4a)

$= \beta y^* + \Gamma x$ (4b)

Thus the transformation to REID and GEID systems does not change the parameters β, Γ, nor the reduced form RF; and in REID and GEID systems SF and RF have the same residuals ε.

3.1. The FP estimation algorithm is the same for REID and GEID systems. The FP algorithm is iterative, say with steps $s = 1, 2, \ldots$ Let ($'$), ($''$) mark the estimation proxies obtained in steps s and $s+1$, respectively, As the algorithm converges the limiting B'', G'', y'' are the FP estimates of β, Γ, y^*.

The FP algorithm has two substeps, alternating between (3a) and (4):

First substep. Multiple OLS regression of y on y' and x gives B'' and G'';

$y = B'' y' + G'' x + e''$ (5)

Second substep. Using B'' and G'' from the first substep, y'' is obtained from (4):

$$y'' = B'' y' + G'' x \qquad (6a)$$

or alternatively,

$$y'' = [I - B'']^{-1} G'' x \qquad (6b)$$

Starting values, $s = 1$. Experience shows that different starting values, e.g.

$$y' = 0; \quad y' = y; \quad y' = \text{the first stage of TSLS} \qquad (7)$$

in most cases lead to the same FP estimates B, G, y^*.

3.2. The FP algorithm (5)-(6) has been improved, generalized and applied in numerous contributions, including the subsequent #1 - #5; see H. Wold, ed. 1980.

*1 The SF may involve identities with specified parameters B, G; E.Lyttkens, 1953.

*2 The convergence of the FP algorithm may be improved by a relaxation parameter ; A. Ågren, 1972.

*3 RFP (Recursive FP) estimation: The two substeps (5)-(6) are performed recursively, without delay in using the flow of new information; L. Bodin, 1974.

*4 To speed up the flow of recursive information in *3 the SF is reordered so as to minimize the number of parameters $b_{ik} \neq 0$ with $k > i$; L. Bodin, 1974.

*5 Applications to real-world models and data, as well as to simulated data, and comparisons with TSLS, LIML and other methods for estimation of IS systems; R. Bergström, 1974.

3.3 The FP method continues to develop, as seen from two books published this year. R.Bergström & H.Wold, 1983, report FP analysis of a large Polish model, and subject FP models to LS methods of model evaluation. Dr. M. Lösch, July 1983, gives a thorough review of FP modeling, and extends the method to the estimation of Rational Expectations (RE) models. Bergström concludes that FP gives predictions that by a quantum leap are more accurate than TSLS and other methods. Lösch lauds the simplicity and speed of FP estimation, as well as the accuracy of the ensuing predictions.

4. Model building: The ML and LS approaches

ML (Maximum Likelihood) methods are the mainstream of contemporary statistics and econometrics. The FP and PLS methods are LS (Least Squares) methods, and therefore I must discuss their reach and limitation relative to ML methods. The comparison is of special importance in the context of models for large complex systems.

The ML methods of statistical inference have a general and well elaborated framework for (i) ML estimation, (ii) hypothesis testing, and (iii) standard errors (SEs) for the estimated parameters. For the LS methods a counterpart to (i-iii) has emerged, namely (j) LS estimation, (jj) Stone-Geisser (SG) testing for predictive relevance, and (jjj) SE assessment by John Tukey's jackknife.

Although (jj-jjj) are fundamental tools of model building they are as yet largely unknown. Hence the next part of my talk will be a discourse on ML vs. LS modeling, with emphasis on fundamental features of assumptions, parameter estimation, model evaluation, and assessment of SEs.

4.1 Assumptions

ML (Maximum Likelihood)	LS (Least Squares)
(a) The observations are jointly ruled by a specified multivariate distribution;	(a) LS is distribution-free, except for predictor specification;
(b) the distribution is subject to independent observations.	(b) independence of the observations is not required,

The LS assumptions are more general by a quantum leap. In consequence, LS is of more broad scope than ML both in theoretical and applied work; LS is more flexible in the adaptation to specific features in the applications.

4.2 Parameter estimation

The Likelihood Function is maximized.	The residual variances are minimized.

In the special case of controlled experiments with nonrandom stimulus variables, ML and LS parameter estimates are numerically the same. Otherwise, the ML and LS estimation methods give more or less different results. Under general regularity conditions, ML parameter estimates have optimal accuracy, and LS predictions have optimal accuracy. Hence there is a choice between parameter and prediction accuracy; in general you cannot have both.

ML estimation is technically difficult, and the difficulties increase with the size of the model. In comparison, the implementation of LS estimation is easy, and the size of the model is rarely a problem. As a rule, LS is speedy on the computer.

4.3. Model evaluation

Hypothesis testing by the Likelihood Ratio	The SG test for predictive relevance

Every model is an approximation, a more or less close approximation. Hence the yes-or-no question of ML hypothesis testing is wrongly posed. The ML null hypothesis is that the model is true; as is well known the Likelihood Ratio will reject the model sooner or later as N (the number of observations) increases. The SG test criterion Q^2 is an R^2 evaluated without lose of degrees of freedom. If $Q^2 < 0$ the model is not predictive, whereas $Q^2 > 0$ indicates the degree to which the model is predictive.

4.4 Standard errors (SEs)

SE assessment by the classical formula	SE assessment by Tukey's jackknife. The SG test gives jackknife SEs as a by-product.
On the ML assumptions, and asymptotically for large N, the classical SE is the smallest possible. The classical SE is robust w.r.t. the distributional assumption (a), but not w.r.t. the independence assumption (b).	Jackknife SEs are realistic assessments, in virtue of the distribution-free LS assumptions (a)-(b). Experience shows that classical SEs typically are underestimated, often by 50%, 100%, or more.

4.5 The distance between substantive theory and statistical technique in scientific modeling.

As illustrated in Fg.2 the FP and PLS estimation methods reduce the distance between substantive theory and statistical technique. Causal Chain systems are seen as the prototype model of general scope for causal-predictive analysis.

For a Causal Chain model to be realistic and useful in applied work the various relations of the SF must be realistic. OLS regression provides consistent parameter estimation of Causal Chain systems, and thanks to the simplicity of OLS this is a substantial advantage of Causal Chains. In the passage to large complex systems, however, it becomes increasingly difficult to design Causal Chain systems that are realistic in all detail, and so the need arises to simplify the model design.

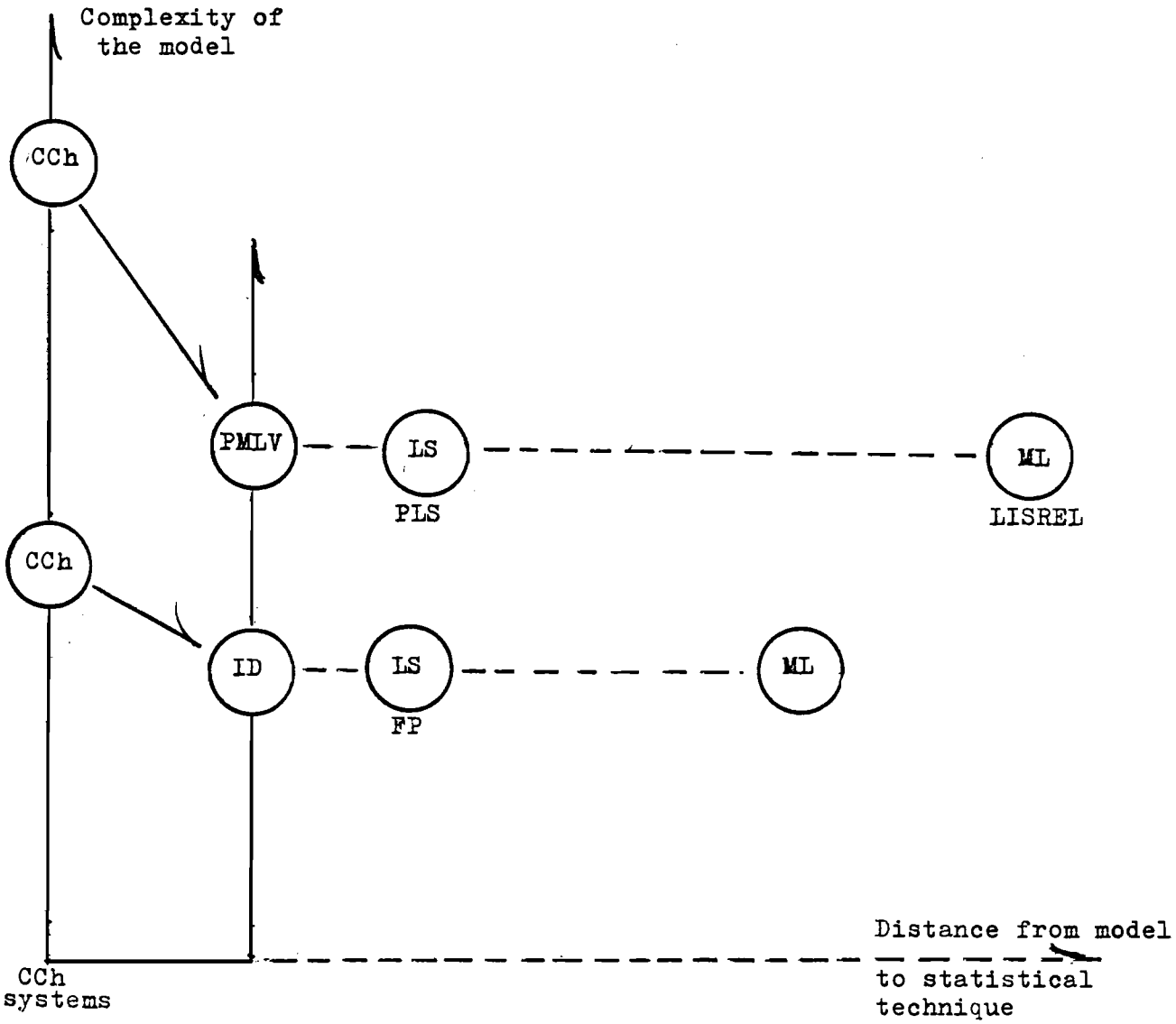

Figure 2. ID (Interdependent) systems and PMLVs (Path Models with Latent Variables) seen as simplifications of CCh (Causal Chain) systems. The distance from substantive theory to statistical technique is larger for ML (Maximum Likelihood) than for LS (Least Squares) estimation.

As illustrated in Fg.2 the simplification in the theoretical design relative to the corresponding Causal Chain system is larger for PMLVs than for ID systems. Fg.2 further illustrates that ML estimation is technically difficult both for PMLVs and ID systems, that LS reduces the distance from the theoretical model to the statistical technique, and that the reduction is larger for PMLVs than for ID systems.

In the published reports on ML estimation of PMLVs and ID systems the models are small or smallish, whereas the size of the model is rarely a problem in FP and PLS applications.

5. Modeling with latent variables (LVs)

5.1. Turning from the left to the right in Fg.1, the historical evolution of the simple model I* is illuminating. C.L. Spearman in 1904 launched the General Factor model of human abilities, in symbols:

$$x_{jn} = \pi_j \xi_n + \varepsilon_{jn}, \quad \bar{x}_j = 0, \quad j = 1, J; \quad n = 1, N \qquad (8\,a\text{-}b)$$

where x_{jn} measures the j-th ability of the n-th person, with measurements given as deviations from the means; π_j is the loading of the j-th ability; ξ is the General Factor, and ξ_n the factor score of the n-th person; ε_j is the specific factor of the j-th ability. The specific factors are assumed to be uncorrelated, giving

$$r_{ij} = r(\varepsilon_i, \varepsilon_j) = 0, \quad i \neq j; \quad i, j = 1, J \qquad (9)$$

Since π_j and ξ_n are unknown, a standardization of scales for unambiguity (SSU) is needed; SSU can be achieved by giving the General Factor unit variance:

$$E[\xi - E(\xi)]^2 = 1 \qquad (10)$$

On condition (9) the tetrachoric correlations of the General Factor model are zero:

$$\begin{vmatrix} r_{ip} & r_{iq} \\ r_{jp} & r_{jq} \end{vmatrix} = 0; \quad i, j, p, q = 1, J \qquad (11)$$

which for a long time was the standard test for the validity of the General Factor model (8).

5.2. It was not until the 1920's that the General Factor model was generalized, perhaps partly because there was no general estimation method that honoured the noncorrelation (9) of the specific factors. The Multiple-factor model of L.L.Thurstone 1935, 1947, say with 3 factors, reads:

$$x_{jn} = \pi_{j1} \xi_{1n} + \pi_{j2} \xi_{2n} + \pi_{j3} \xi_{3n} + \varepsilon_{jn}, \quad j = 1, J; \quad n = 1, N \qquad (12)$$

where the general and specific factors are uncorrelated,

$$r(\xi_p, \xi_q) = r(\xi_p, \varepsilon_j) = r(\varepsilon_i, \varepsilon_j) = 0, \quad p, q = 1, 3; \quad i, j = 1, J \qquad (13)$$

The matrix $[r_{ij}]$ is called the correlation structure of the data x_{jn}. Extending the SSU standardization (10), and writing $\Delta_j = \mathrm{var}(\varepsilon_j)$, the Multiple-factor model of the structure is:

$$r_{ij} = \sum (\pi_{iq} \pi_{jq}), \quad r_{ii} = \sum (\pi_{iq}^2) + \Delta_j \; ; \quad i \neq j; \quad i, j = 1, J \qquad (14\,a\text{-}b)$$

In words, the loadings model the correlation structure, except that the specific factor variances must be added in the diagonal.

In lack of estimation methods in keeping with the noncorrelation (9) of the specific factors, the factor models (8) and (12) were often approximated by the corresponding Principal Components models, say in the case (12):

$$x_{jn} = p_{j1} X_{1n} + p_{j2} X_{2n} + p_{j3} X_{3n} + e_{jn} \qquad (15)$$

using the algebraic method of eigenvalues and eigenvectors to estimate the loadings p_{jq} and component scores X_{qn}.

6. Path models with LVs

To quote C. Fornell (1982), a second generation of multivariate analysis emerged in the mid-1960's. An iterative procedure introduced by H. Wold (1966) for estimation of Principal Components models gives loadings and component scores that are numerically equivalent to those given by the algebraic method in terms of eigenvalues and eigenvectors. K.G. Jöreskog in 1967 was the first to give a general estimation procedure for General and Multiple-factor models with uncorrelated specific factors.
In these innovations there is a twofold parting of the ways: difference in purpose, and difference in estimation technique.

Purpose: The General and Multiple-factor models estimate the correlation structure. The Principal Components model (15), also known as Single Value Decomposition, models the data x_{jn} in terms of the estimated loadings and component scores.

Method: Jöreskog uses ML estimation to model the covariance structure. The algebraic estimation of Principal Components is an LS method, and so is Wold's iterative algorithm.

The properties of the ensuing estimates are in line with the general theory of ML and LS estimation. Jöreskog's algorithm gives consistent estimates for the parameters, i.e. for the loadings and the specific factor variances, whereas no estimates are obtained for the factor scores. Wold's Principal Components model gives LS predictions (15) for the data x_{jn}, predictions with minimum variance for the prediction errors e_{jn}, whereas the estimates of the loadings and the component scores in general are **consistent only in a qualified sense (see Sn. 6.1).**

6.1. A key feature of **the LS** estimates of the component scores is that they are weighted aggregates of the data; thus for the first component:

$$X_{1n} = f_1 \sum_j (p_{j1} x_{jn}) \qquad (16)$$

where f_1 is a scalar that gives X_1 unit variance, in accordance with (10). Under mild supplementary conditions the estimates of component scores and loadings are consistent at large; that is, if J (the number of observables) is allowed to increase X_{1n} will in the limit tend to the theoretical value ξ_{1n}, and similarly for the loadings p_{j1}.

6.2 Jöreskog in 1973 extended his ML estimation of Factor models to general path models with LVs, an extension from Model I* in Fg. 1 to Models II* - V*. In 1966 I had given an iterative LS estimation of Canonical Correlations, Models II* in Fg.1. When seeing Jöreskog's ML estimation of Models III*-V* this gave me the clue to extend my LS estimation of Models I*-II* to Models III*-IV*. I shall now give a brief exposition of this device, called Partial Least Squares. Spelled out for Model IV*, the exposition covers Models I*-III* and II-IV as special cases.

7. The basic design of PLS modeling.

*1 The arrow scheme constitutes the theoretical-conceptual design of the model. The investigator is free to design the arrow scheme in accordance with the purpose of the model, his prior knowledge and intuition, and the available data. The arrow scheme specifies the LVs, the "inner" relations between the LVs, for each LV a number of MVs, called indicators, for its indirect observation. The inner relations are the core of the model.

Three introductory examples with endogenous LVs η and exogenous LVs ξ :

Model II*. Structure ξ and Response η , indirectly observed by indicators x_h, y_k.

Model III*. Students' achievements η influenced by Parents' home ξ_1 and School conditions ξ_2.

Model IV*. Adelman's model, 1975: Economic levels ξ_1 and Social conditions ξ_2 influence Political condition η_1 . Economic growth η_2 is influenced by ξ_1, ξ_2 and η_1.

*2 Formal definition of the model

Endogenous LVs and MVs, observed over T cases:

$$\eta_{it} \qquad y_{iht} \qquad i = 1, \ldots, n\ ; \quad h = 1, H_i'; \quad t = 1, T \qquad (17\ a\text{-}b)$$

Exogenous LVs and MVs:

$$\xi_{jt} \qquad x_{jkt} \qquad j = 1, \ldots, m; \quad k = 1, H_k''; \quad t = 1, T \qquad (18\ a\text{-}b)$$

The ranges of the subscripts will often be tacitly understood. Note that t,T in (17)-(18) correspond to n, N in (8).

The cross products of the raw data (17 b)-(18 b) constitute the product data.

*3 Inner relations. Linear relations subject to predictor specification:

$$\eta_t = \beta \eta_t + \Gamma \xi_t + \upsilon_t \qquad (19)$$

$$E(\eta_t | \eta_t, \xi_t) = \beta \eta_t + \Gamma \xi_t \qquad (19a)$$

Two LVs are called adjoint if they are directly connected by an arrow in the arrow scheme.

*4 Outer relations. Linear relations subject to predictor specification:

$$y_{iht} = \pi_{1ih0} + \pi_{1ih} \eta_{it} + \varepsilon_{1iht} \qquad (20)$$

$$x_{jkt} = \pi_{2jk0} + \pi_{2jk} \xi_{jt} + \varepsilon_{2jkt} \qquad (21)$$

$$E(y_{iht} | \eta_{it}) = \pi_{1ih0} + \pi_{1ih} \eta_{it} \qquad (20a)$$

*5 Prediction by the model. The inner relations (19) predict endogenous LVs in terms of endogenous and exogenous LVs. The outer relations (20) predict endogenous indicators in terms of their LV.

Substitutive prediction: Endogenous LVs and MVs predicted in terms of endogenous and/or exogenous LVs. 1) In the inner relation for an LV in (19) one explanatory endogenous LV is substituted in terms of its inner relation. 2) In the outer relation (10) for an endogenous indicator the LV is substituted in terms of its inner relation.

Repeated substitution: Repeated use of 1)-2) to substitute one or more endogenous LVs in terms of their inner relations.

7.1 The model in estimated form:

Latent variables:

$$Y_{it} = \text{est}(\eta_{it}) = \sum_h (w_{1ih}\, y_{iht}); \quad X_{jt} = \text{est}(\xi_{jt}) = \sum_k (w_{2jk}\, x_{jkt}) \qquad (22\ a\text{-}b)$$

The weights will be determined in the PLS algorithm. They are auxiliary parameters that do not belong to the formal model.

Inner relations:

$$Y_t = B\, Y_t + G\, X_t + u_t$$

Outer relations:

$$y_{iht} = p_{1ih0} + p_{1ih}\, y_{it} + e_{1iht}; \quad x_{jkt} = p_{2jk0} + p_{2jk}\, x_{jt} + e_{2jkt} \qquad (23\ a\text{-}b)$$

7.2 PLS estimation of the model.

The algorithm will be set forth with raw data input. Without loss of generality we take the raw data to be measured as deviations from their means, which gives

$$\bar{y}_{ih} = \bar{x}_{jk} = 0 \qquad i = 1, n;\ i = 1, m;\ h = 1, H_i';\ k = 1, H_k'' \qquad (24)$$

*1 First stage. The first stage estimates weights v that are proportional to w:

$$Y_{it} = \sum_h (w_{1ih}\, y_{iht}) = f_{1i} \sum_h (v_{1ih}\, y_{iht}); \quad X_{jt} = \sum_k (w_{2jk}\, x_{jkt}) \qquad (25\ a\text{-}b)$$

where f_{1i} is a scalar that gives Y_i unit variance, and similarly for X_{jt}. The weights are determined by weight relations. Unifying the notation for the LVs by $\zeta = (\eta, \xi)$, the weight relations for any LV, say ζ_a, involve a sign weighted sum, denoted SwS_{at}, of estimates of those LVs that are adjoint to ζ_a, say $\zeta_{a'}$:

$$SwS_{at} = \sum_{a'} [(\pm)_{aa'} Z_{at}] \qquad (26a)$$

with

$$(\pm)_{aa'} = \text{signum}\ r(Z_a, Z_{a'}) \qquad (26b)$$

For each ζ_a the investigator has the option to choose between two types of weight relations, called Modes A and B, which take the form of simple and multiple OLS regressions of z_{ah} on SwS_a:

Mode A. For each h the simple OLS regression of z_{ah} on SwS_a:

$$z_{aht} = v_{ah}\, SWS_{at} + d_{aht} \qquad (27\ a)$$

Mode B. The multiple OLS regression of SwS_a on z_{ah}:

$$SwS_{at} = \sum_h (v_{ah}\, z_{aht}) + d_{at} \qquad (27\ b)$$

As a rule of thumb, Mode A should be selected for the endogenous LVs, Mode B for the exogenous LVs. However, Mode A should be selected for exogenous LVs ζ_j with many indicators, say $H''_j > T/10$.

The PLS algorithm alternates between (25) and (27), and proceeds in steps $s = 1, 2, \ldots$ with substeps for the various LVs. In analogy to RFP estimation each substep uses LV proxies obtained in earlier substeps.

The starting values in step $s=1$ are largely arbitrary, say $v_{ah} = 1$ for all a, h.

Passage to the limit. The iterative procedure continues until each estimated w_{at} converges according to a conventional stopping rule, say:

$$(w^{s+1}_{at} - w^{s}_{at})/w^{s}_{at} \leq 10^{-5} \qquad (28)$$

for all a and t.

*2 Second stage of the PLS algorithm. Using the LVs estimated in Stage 1, the noniterative second stage estimates the inner and outer relations by OLS regressions. The estimated relations take the form (19) - (21), with zero location parameters because of the zero means (24).

Dropping (24), it is immediate matter to estimate the location parameters, as always in LS estimation. Thus for the LVs and the outer relations:

$$\bar{z}_a = \sum_h (w_{ah}\, \bar{z}_{ah}); \qquad z_{ah0} = \bar{z}_{ah} - p_{ah}\, \bar{z}_a \qquad (29)$$

Prediction and Substitutive prediction. The theoretical predictions 6,*5 carry over to the estimated model. Furthermore, substituting the estimated LVs by the weighted aggregates (25 a-b), the model gives predictions of endogenous LVs and MVs in terms of MVs

*3 Model evaluation. The SG test for predictive relevance and the jackknife assessment of standard errors are of general scope in LS modeling, and are part of the basic PLS design; cf. 4.3 - 4.4.

The power of classical model evaluation rests on the aggregation over the case values ($t = 1, T$). PLS modeling involves a twofold aggregation, over the case values and for each LV over its indicators, and the aggregation (25) ver the indicators adds to the power of the model estimation. For example, in a real-world model with $T = 10$ and two exogenous LVs with in all 27 indicators, the SG test gave $Q^2 = .44$, indicating that the model is predictive; H. Wold (1978, 1980, 1983a).

*4 Product data input. It is immediate matter to carry over the PLS algorithm (24)-(29) from raw data input to product data input. The ensuing procedure is more speedy on the computer, and the resulting parameter estimates are numerically the same, except for rounding errors. The difference is that for each LV the product data only give aggregate values over t, such as means and variances of the LV estimates, whereas raw data input is needed to obtain estimates of the case values Z_{at} ($t = 1, T$) for each LV.

*5 Computer programs of the PLS algorithm (manual and tapes) are available at nominal cost, and cover both raw data input and product data input; Lohmöller, 1981.

8. PLS modeling: Extensions of the basic design.

Thanks to the generality of being distribution-free, PLS modeling allows an array of extensions. In the extensions *1- *5 the requisite changes in the basic PLS algorithm are direct matter. In *6 - *7 the generalizations are more advanced.

*1 The indicators of any LV can be replaced or supplemented by some functional transform, such as

$$\log x_{jk} \ , \quad x_{jk}^2 \ , \quad x_{jh}\, x_{jk} \ , \ \ldots \tag{30}$$

*2 In the inner relations one or more LVs may be replaced by MVs.

*3 Categorical variables and contingency tables. Primarily designed for scalar variables, the PLS algorithm carries over to categorical indicators and thereby to contingency tables. A contingency table with just one categorical variable in each margin is the covariance matrix of the two variables.

*4 Higher dimensions of the LVs. The basic PLS design estimates the first dimension of each LV. Higher dimensions of an LV can be estimated consecutively, using as data input the residuals of the outer relations.

Special cases: Principal components and Canonical correlations of higher "orders", in PLS called "dimensions."

Speaking broadly, if PLS estimation gives markedly different weights for an LV when using Mode A and Mode B, this is an indication that this LV has more than one dimension.

*5 Hierarchic structure of the LVs as modeled by PLS is analogous to the hierarchic structure in psychometric factor analysis pioneered by L.L.Thurstone (1935, 1947).

6 Models with feedbacks or interdependencies in the inner relations: Model V in Fg.1. The estimation combines the FP algorithm with the second stage of the PLS algorithm.

*7 Nonlinearities in the inner relations. A rather straightforward case:

$$\eta_t = \beta_0 + \beta_1 \xi_t + \beta_2 \xi_t^2 + \nu_t \tag{31}$$

A sophisticated case at the research frontier of PLS: The three LVs of the model $\xi_{1t}, \xi_{2t}, \eta_t$ form the first level of a hierarchic structure where at the second level they satisfy a third-degree eqtion:

$$\eta_t^3 + \xi_{1t}\, \eta_t + \xi_{2t} = 0 \tag{32}$$

In this model η_t may involve discontinuities in the sense of Thom's catastrophe theory.

9. Applications of PLS modeling.

Initiated some ten years ago, PLS is now firmly consolidated, and is rapidly gaining momentum. The reported applications range from reproducible data in natural science and medicine to the *nonreproducible* data of socioeconomic, behavioural and political sciences. The central and yet broad realm of PLS is reserch contexts that simultaneously are data-rich and theory-primitive. Ever larger models are being reported, and it is safe to say that PLS has its forte in the analysis of large complex systems.

*1 One of the largest PLS models analyzes an educational system; R. Noonan & H. Wold, 1983. 191 MVs are grouped as indicators of 59 LVs; there are 32 inner relations,

including 14 hierarchic structures; 1,300,000 cells (1.3 megabyte) are required in the memory space; the PLS estimation converges in 4 iterations, using a total of 44.6 seconds computer time.

*2 In substantive research, PLS has inspired investigators to construct large models. For example, the educational model in *1 is a synthesis of six recent models that deal with specific aspects of educational systems.

*3 The broad scope and flexibility of the PLS approach is reflected in the diversity of PLS applications. The data can be scalar, ordinal, or categorical; the inner relations can be linear or nonlinear; the purpose can be prediction, classification, or causal analysis.

*4 The SIMCA program for classification, S. Wold (1978) is based on disjoint Principal Components models estimated by PLS, where the appropriate number of dimensions is assessed by the Stone-Geisser test for predictive relevance.

8.1 Opinions among investigators using PLS. To summarize from discussions with the investigators, the following advantages of PLS have been emphasized.

*1 The broad scope and flexibility of the PLS approach in theory and practice. The conceptual-theoretical definition of a PLS model is given by its arrow scheme, which suffices as a basis for the formal specification of the model and for the PLS algorithm.

*2 "Instant estimation." PLS is a rapid affair, even if the model is large; cf. 8,*1.

*3 Parsimony. In a PLS model with J LVs there are TJ case values of the LVs, $\sum_j(H_j)$ loadings, say n_b inner parameters, and $J + \sum_j(H_j)$ location parameters. To estimate these unknowns PLS is parsimonous in using weights w_{jh} as auxiliary tools, namely in all $\sum_j(H_j)$ weights.

*4 PLS has reduced the distance between statistical theory and substantive analysis. A PLS model devlops by a dialogue between the investigator and the computer. Tentative improvements of the model - such as the introduction of a new LV, an indicator, or an inner relation, or the omission of such an element - are tested for predictive relevance by the SG test. The various pilot studies are a speedy and low cost matter.

9. Outlook. Breaking away from the ML mainstream, and placing emphasis on applied work, PLS has from the outset attracted active interest from substantive researchers. PLS modeling combined with the SG and jackknife methods now constitutes a distribution-free approach of general scope for quantitative systems analysis.

In this broad perspective PLS modeling is at an early stage of evolution. There is an abundance of potential applications, including many fields where systems analysis is still at the qualitative stage. In the passage from qualitative to quantitative analysis, of course, the scarcity or lack of adequate data is the main problem. Hence for a long time to come the progression of PLS to new fields of quantitative systems analysis will very much be a matter of data work, substantive theory and data work.

BIBLIOGRAPHY

	Sns. where the reference is of specific import
Bergström, R., Wold, H. (1983) Fix-Point estimation in theory and practice, Vol.23, Applied Statistics and Econometrics, eds. G.Tintner, H.Strecker and E. Féron. Göttingen: Vandenhoeck and Ruprecht.	3.3; 4.3; 4.4
Fornell, C. (ed. 1982) A Second Generation of Multivariate Analysis, Vol.1 - Methods. New York: Praeger.	6
Jöreskog, K.G., Wold, H. (eds. 1982) Systems under Indirect Observation: Causality * Structure * Prediction (2 vols.). Amsterdam: North-Holland.	4; 5; 6
Kappel, R., Schwarz, I.A. (1981) Systemforschung 1970-1980, Entwicklungen in der Bundesrepublik Deutschland. Göttingen: Vandenhoeck and Ruprecht.	1
Lohmöller, J.,B. (1983) Path Models with Latent Variables and Partial Least Squares (PLS) estimation, Doctoral Dissertation, München, Hochschule der Bundeswehr.	5; 6; 7; 7.2:*5
Lösch, M. (1983) Fixpunkt-(FP-)Schätzverfahren für Modelle mit rationalen Erwartungen (RE-Modelle), Habilitation Thesis, Hochschule St.Gallen.	3.3
Noonan, R., Wold, H. (1983) Evaluating School Systems Using Partial Least Squares, Vol. , Evaluation in Education: An International Review Series, eds. B. Choppin and T.N. Postlethwaite. Oxford: Pergamon. (In press.)	9:*1
Wold, H. (ed. 1980) The Fix-Point Approach to Interdependent Systems. Amsterdam: North-Holland.	1; 2; 3
Wold, H. (1982) Soft Modeling: The basic design and some extensions, in Jöreskog and Wold, eds. 1982, Vol.2: 1-54.	4; 5; 6; 7; 8; 8:*7 (31); 9
Wold, H. (1983a) Fix-Point Method, in Encyclopedia of Statistical Sciences, eds. S. Kotz & N.L. Johnson, 3: 148-156.	2; 3; 4.3; 4.4
Wold, H. (1983b) Systems analysis by Partial Least Squares, NATO Advanced Research Workshop on Analysis of Qualitative Spatial Data, Amsterdam, March 28 - April 1, 1983.	4; 7.2:*3
Wold, H., Bertholet, J.-L. (1982) The PLS (Partial Least Squares) approach to multidimensional contingency tables, Metron 40, 1-2: 303-326.	4.3; 8:*3
Wold, H., Mensch, G.O. (1983) Nonlinear extensions in soft modeling, WSOM Workshop Paper 83-017, Case Western Reserve University, Cleveland, Ohio.	8:*7 (32)
Wold, S. (1978) Cross-validatory estimation of the number of components in factor and principal components models, Technometrics 20: 397-405.	4.3; 9:*4