MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

SOFT MODELLING: INTERMEDIATE BETWEEN TRADITIONAL MODEL BUILDING AND DATA ANALYSIS

HERMAN WOLD

*University of Uppsala, Department of Statistic, Uppsala, Sweden**

0. Introduction

NIPALS (*Nonlinear Iterative Partial Least Squares*) soft models are primarily intended for interdisciplinary problems, or more generally for complex situations where prior information is scarce. Blocks of manifest (directly observed) variables are the structural units of a NIPALS soft model. The data may be observations over time or a cross-section. Each block of observables serve as indicators for a latent (indirectly observed) variable, estimated as a weighted aggregate of its indicators. A typical application: School pupils' achievement as influenced by Home background and School conditions: see Fig. 1.

Well-known problems of multivariate analysis are (i) how to define "quality of life", "consumer sentiment", "home background", and other "soft" notions by the way of suitably weighted aggregates of a block of indicator variables, and (ii) how to assess the inner structure of a block of observables using the family of methods known as "multidimensional scaling". While the current methods under (i)–(ii) are concerned with just one block of observables, NIPALS soft modelling in general is a multi-block, causal-predictive approach. The arrow scheme of a soft model defines the conceptual design of the model, showing the hypothetical "inner" relations between the latent variables. The inner relations are assumed to form a causal chain system. The weights used in estimating a latent variable are determined by a system of weight relations. This is an auxiliary system that extracts information from the latent variable to be estimated and its adjacent variables; that is, those latent variables with which it is directly connected in the arrow scheme. For each latent variable and its indicators the investigator has the option to choose between two versions of the weight relations, called *Modes A* and *B*. The weight relations are called *Mode C* if each of Modes A and B is chosen at least once in the model.

The first principal component and the first canonical correlation can be seen as special cases of NIPALS soft models, with one or two latent variables, and with

* *pro tem* University of Geneva.

weight relations Mode A and B, respectively. The NIPALS estimation algorithm yields the indicator weights, and thereby explicit estimates for the case values of the latent variables, and then in turn the estimates for parameters, residuals, and other unknowns in the model.

Generalization aspects: (a) Latent variables in two or more dimensions; (b) Inner relations that form an interdependent system.

1. NIPALS soft models: Principles of specification and estimation

The principles will be briefly set forth using Models C331 and C332 for illustration of the general design. For further details, including references to earlier papers, see [25]–[27], [30*].

1.1. *Conceptual design of NIPALS soft models: The arrow scheme*

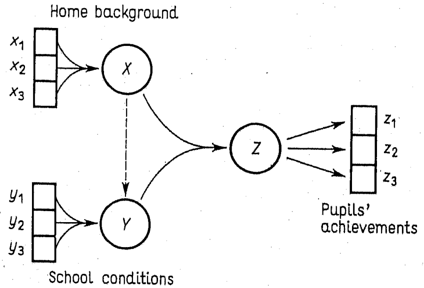

Fig. 1. Arrow scheme for NIPALS soft models C331 (without broken arrow) and C332 (broken arrow included) as applied to explain School pupils' achievements as influenced by Home background and School conditions

The arrow scheme shows how a complex problem is structured and simplified in NIPALS soft modelling. Observables are grouped into blocks, and taken as indicators for as many latent variables. Each latent variable is estimated by a weighted aggregate of its indicators. The code letters A, B, C indicate which mode of NIPALS estimation the investigator has chosen.

The inner relations between the latent variables are the causal-predictive core of the model. All information between the blocks of a soft model is conveyed by the latent variables via the inner relations. No information between the blocks is directly conveyed by the indicators, nor by the residuals that occur in the model.

1.1.1. *Notations in the arrow scheme.*

Squares: Manifest variables, grouped as indicators into blocks.

Circles: Latent variables.

Arrows between latent variables: Inner relations.

Arrows between a latent variable and its indicators mark the choice of weight relations: Single arrows for Mode A, a bundle of arrows for Mode B.

Arrow heads: Residuals of weight relations and inner relations. Residuals are not indicated otherwise in the arrow scheme.

1.1.2. Model C331, as shown in the arrow scheme, and as the code indicates, has three blocks of indicators, three latent variables, and one inner relation.

1.1.4. The observables are grouped into blocks so as to be indicators for holistic, "soft" concepts; in the present case Students' achievements, Home background, and School conditions. As indicators for a "soft" concept the observables in one and the same block usually are correlated, and more correlated than indicators in different blocks.

1.2. *Operative specification of NIPALS soft models*

The exposition in 1.2–1.4 covers the prototype case of NIPALS soft models when the structural relations are linear, and the model is specified in terms of the data.

1.2.1. *Data.* In Model C331 the three blocks of indicators will be denoted by

$$\text{(1a–c)} \qquad x_h (h = 1, \ldots, H), \qquad y_j (j = 1, \ldots, J), \qquad z_k (k = 1, \ldots, K).$$

(i) The data may be given as *raw data*, say N cases (time series or cross section data), to be denoted by x_{hn}, y_{jn}, z_{kn} $(n = 1, \ldots, N)$.

(ii) *Moment data.* In this case the data is the first and second order moments of the raw data, namely the vector of averages $L = (\bar{x}_h, \bar{y}_j, \bar{z}_k)$ of the observables, and their dispersion (variance-covariance) matrix, denoted by $R(x, y, z)$ or $V(x, y, z)$ according as the indicators are or are not standardized to unit variance.

1.2.2. *Latent variables.* The three latent variables of the model are estimated by

$$\text{(2a–c)} \qquad X = f_1 \sum_h (w_{1h} x_h), \qquad Y = f_2 \sum_j (w_{2j} y_j), \qquad Z = f_3 \sum_k (w_{3k} z_k),$$

where f_a $(a = 1, 2, 3)$ are standardizing factors that give X, Y, Z unit variance.

(i) With raw data, formulas (2) give case values X_n, Y_n, Z_n $(n = 1, \ldots, N)$ for the latent variables in terms of the weights w_{1h}, w_{2j}, w_{3k} and the case values 1.2.1 (i) of the indicators.

(ii) With moment data, elementary operations on (2) give the averages $\bar{X}$, $\bar{Y}$, $\bar{Z}$ and the correlation matrix $R(X, Y, Z)$ of the latent variables, as well as the covariance matrix of the latent variables and the indicators, all through in terms of the weights w_{1h}, w_{2j}, w_{3k} and the moment data 1.2.1. (ii).

1.2.3. *Block structure.* This is the set of simple OLS regressions of any indicator on its latent variables,

$$\text{(3a-c)} \qquad x_{hn} = p_{1h0} + p_{1h} X_n + u_{1n}; \quad y_{jn} = p_{2j0} + p_{2j} Y_n + u_{2j}; \quad z_{kn} = p_{3k0} + p_{3k} Z_n + u_{3n}.$$

The coefficients p_{1h}, p_{2j}, p_{3k} are called the *loadings* of the three blocks of indicators.

Since both the loadings and the latent variables are unknown, some standardization of scales is necessary to avoid ambiguity in the model. In soft modelling, all latent variables are standardized to unit variance; cf. 1.2.2.

For each indicator the investigator should specify a priori for each loading whether its sign is expected to be positive or negative. Soft modelling provides two complementary measures for the relevance of an indicator: loading and weight.

1.2.4. *Inner relations*. The inner structural relation of Model C331 explains the third latent variable in terms of the two first ones,

$$Z = b_{30}+b_{31}X+b_{32}Y+e_3. \tag{3a}$$

1.2.5. *Weight relations*. For the three blocks of Model C331 the weight relations are specified as follows, the location parameters being omitted:

$$\begin{aligned} &x\text{-block:} \quad Z = \sum_h (w_{1h}x_h)+d_1; \\ &y\text{-block:} \quad Z = \sum_j (w_{2h}y_j)+d_2; \end{aligned} \tag{4a-b}$$

$$z\text{-block:} \quad z_k = w_{3k}(b_{31}X+b_{32}Y)+d_{3k} \quad (k = 1, \dots, K). \tag{4c}$$

The rationale of the specification (4a–c) will be explained in 1.3.3. In passing we note that in accordance with 1.1.1 and the arrow scheme, the weight relation for each of the x and y-blocks is a multiple regression, for the z-block a set of simple regressions.

1.3. *The design of NIPALS soft models*: *Four fundamental principles*

1.3.1. *The reduction principle*. NIPALS soft models are designed so that simple models form special cases of more complex models. For example, if the y-block is deleted in Model C331, formulas (1)–(4) will reduce to the specification for Model C221, also denoted by C2. If also the x-block is deleted, Model C2 will reduce to Model A110, or briefly A1.

(i) NIPALS models A110 and B221 are numerically equivalent to the first principal component and the first canonical correlation, respectively; [13], [14], [24].

1.3.2. *Estimation as a corollary of specification*. In NIPALS soft modelling the arrow scheme provides as an immediate corollary the algorithm for NIPALS estimation of the unknowns of the model. The NIPALS algorithm proceeds in three stages. The two first use the indicators as standardized to zero mean; the third stage estimates the location parameters.

The first stage of the NIPALS algorithm is an iterative estimation of the latent variables as weighted aggregates of their indicators, with weights determined by the weight relations.

1.3.3. *For the design of the weight relations the investigator can choose between two versions of the least squares principle*. For each latent variable the weight relations are designed to maximize the correlation of the latent variable and a linear form of those latent variables, called adjacent, with which it is directly connected in the arrow scheme. The investigator havin gthe option to choose between the weight relations Mode A or B, these make different twists of the least squares principle.

Weight relations, Mode A: The weight of any indicator is the simple OLS regression coefficient of the indicator on a linear form of the adjacent latent variables.

Weight relations, Mode B: The indicator weights are the multiple OLS regression coefficients of the linear form of adjacent latent variables on the block of indicators.

Thus in Mode A the indicators enter as predictands in a set of simple OLS regressions, in Mode B as predictors in a multiple OLS regression. Hence the ensuing estimates of the latent variable as a weighted aggregate of its indicators is called predictand-weighted and predictor-weighted, respectively.

Depending on whether a latent variable is estimated Mode B or A, the estimate is or is not invariant to changes in the scales of its indicators.

1.3.4. *The device of "partial least squares"*. The combined use of several OLS regressions in 1.3.3 constitutes the NIPALS device of "partial least squares". Let the proxies for the unknowns obtained is estimation step s be denoted by superscripts (s). For given proxies $X_n^{(s)}$, $Y_n^{(s)}$, $Z_n^{(s)}$ of the latent variables, the inner relation (3b) determines the proxies $b_{30}^{(s)}$, $b_{31}^{(s)}$, $b_{32}^{(s)}$ so as to minimize the residual variance var $(e_3^{(s)})$. Similarly, for any weight relation, say (4a): for given proxies $Z_n^{(s)}$ the weights $w_{1h}^{(s)}$ $(h = 1, \dots, H)$ are determined so as to minimize the variance of the residual $d_1^{(s)}$, thereby maximizing the correlation between $X_n^{(s)}$ and $Z_n^{(s)}$.

In the limit as $s \to \infty$, various least squares criteria are fulfilled simultaneously, giving estimates for the unknowns that fulfil the operative definition of the model.

(i) Well to note, the "partial least squares" conditions 1.3.4 in general will not imply that the resulting NIPALS estimation will fulfil an overall, total least squares criterion.

1.4. *Estimation of NIPALS soft models, exemplified by Models C331 and C332.*

1.4.1. *Model* C331. We shall spell out the estimation procedure with raw data input. Until further notice we assume that the observables (1) are measured as deviations from their means, implying $\bar{x} = \bar{y} = \bar{z} = 0$ for all h, j, k.

1.4.2. *Start*, $s = 1$. Arbitrary starting values are specified for g_k, say $w_{3k}^{(s)} = 1$ $(k = 1, \dots, K)$. Then (2c) gives first $f_3^{(1)}$ and then $Z_n^{(1)}$ $(n = 1, \dots, N)$.

1.4.3. *General step from* s *to* $s+1$.

(i) Using $Z_n^{(s)}$ $(n = 1, \dots, N)$, the multiple OLS regression (4a) gives $w_{1h}^{(s+1)}$ $(h = 1, \dots, H)$.

(ii a) Using $w_{1h}^{(s+1)}$, standardization of the aggregate (2a) to unit variance gives $f_1^{(s+1)}$.

icm©

(ii b) Using $w_{1h}^{(s+1)}$ and $f_1^{(s+1)}$, the aggregate (2a) gives $X_n^{(s+1)}$ $(n = 1, \ldots, N)$.

(iii)–(iv) As in (i)–(ii, a–b), again using $Z_n^{(s)}$, the regression (4b) gives $w_{2j}^{(s+1)}$ $(j = 1, \ldots, J)$, and the aggregation (2b) gives $f_2^{(s+1)}$ and $Y_n^{(s+1)}$ $(n = 1, \ldots, N)$.

(v) Using $Z_n^{(s)}, X_n^{(s+1)}, Y_n^{(s+1)}$ $(n = 1, \ldots, N)$ the multiple OLS regression (3d) gives $b_{31}^{(s+1)}$ and $b_{32}^{(s+1)}$.

(vi) Using $X_n^{(s+1)}, Y_n^{(s+1)}$ $(n = 1, \ldots, N)$ and $b_{32}^{(s+1)}, b_{32}^{(s+1)}$, the simple OLS regression (4c) with k fixed gives $w_{3k}^{(s+1)}$ $(k = 1, \ldots, K)$.

(vii) Using $w_{3k}^{(s+1)}$ and proceeding as in (i)–(ii, a–b), the the aggregation (2c) gives $f_3^{(s+1)}$ and $Z_n^{(s+1)}$ $(n = 1, \ldots, N)$.

1.4.4. *The limit as* $s \to \infty$. Subject to a suitable convergence criterion, the limiting estimates are denoted as in (1)–(4), for example,

(5a–c) $$w_{1h} = \lim_{s\to\infty} w_{1h}^{(s)}, \quad X_n = \lim_{s\to\infty} X_n^{(s)}, \quad b_{31} = \lim_{s\to\infty} b_{31}^{(s)}.$$

1.4.5. *The block structure.* The second stage of the NIPALS algorithm estimates the block structure 1.2.3, using the latent variables estimated in the first stage.

1.4.6. *Nonzero means.* When the estimation 1.4.2–1.4.5 has been performed, the assumption in 1.4.1 of zero means can be revoked, with no change in the estimates 1.4.4, just as in OLS regression. The following formulas for the ensuing estimates of the location parameters will suffice for illustration,

(6a–c) $$\bar{X} = f_1 \sum_h (w_{1h}\bar{x}_h), \quad p_{1h0} = \bar{x}_h - p_{1h}\bar{X}, \quad b_{30} = \bar{Z} - b_{31}\bar{X} - b_{32}\bar{Y}.$$

1.4.7. *The sign of the standardizing factors.* Since the standardization factors in (2) have ambiguous sign $\pm$, the signs should be chosen so as to make for agreement with the signs of the loadings postulated in 1.2.3. The degree of agreement makes a partial test for the realism of the model; cf. [2].

1.4.8. *Moment data input.* Having spelled out the estimation procedure 1.4.2–1.4.7 for raw data input, it is direct matter to adapt the procedure for moment data input 1.2.1(ii). The estimation output will include the coefficients of weight relations, inner relations, and block structure; the dispersion matrices of latent variables and residuals; and the covariance matrices of observables, latent variables, and residuals.

1.4.9. *Models* C331 *and* C332: *Thresholds in NIPALS soft modelling.*

(i) In the special cases of Models A221, B221, C221, the first stage of the NIPALS algorithm does not estimate the inner relation, which then is estimated in the second stage. For the same two-block models the inner relation can be the sole target of the first estimation stage, a device which leads to numerically the same results; [8], [20].

(ii) The passage from two to three blocks of observables is crucial in NIPALS soft modelling. When coming to C331 and more complex models the two estimation devices referred to in (i) are no longer equivalent and self-contained. Turning next to Model C332, we shall see that the passage from one to two inner relations is equally crucial in introducing a new feature into the estimation procedure.

1.4.10. *Model* C332. As seen from the chart, Model C332 has the same arrow scheme as C331, except that one more inner relation enters, namely for Y in terms of X. For the same data 1.2.1, Models C331 and C332 have the same operative specification (2a–c) for the latent variables, for the inner relation (3b) that explains Z, and for the weight relations (4c) of the z-block. The inner relation for Y in Model C332 is specified by

(7) $$Y = b_{20} + b_{21}X + e_2.$$

The new feature of Model C332 is the weight relations for the x and y-block, (4a–b) being replaced by

(8a) $$W_2 s_{21} b_{21} Y + W_3 s_{31} b_{31} Z = \sum_h (w_{1h} x_h) + d_1,$$

(8b) $$W_2 s_{21} b_{21} X + W_3 s_{32} b_{32} Z = \sum_j (w_{2j} y_j) + d_2.$$

The factors W_2 and W_3 are nonnegative weights that the model builder has the option to attach a priori to the two inner relations,

(9) $$W_2 + W_3 = 1, \quad W_2 > 0, \quad W_3 > 0.$$

The sign factors s_{ik} are designed to avoid that the left-hand terms in (8a–b) cancel when forming the normal equations for the OLS regression coefficients to the right; the special case

(10) $$s_{31} = \mathrm{sign}[b_{31} r(X, Z)]$$

will suffice to indicate the general design of the sign factors.

1.4.11. *Computer programs.* The estimation procedure 1.4.1–1.4.7 with options for raw data or matrix data input has been programmed for the computer by Areskoug [6], covering A, B and C-models with one inner relation, and up to three blocks of observables; Hui-Hausman [11] have made a program for up to twelve blocks of observables, and otherwise of the same scope. Apel [5] has programmed the general procedure 1.4.1–1.4.10 for raw data input and up to five blocks of observables, with any design for the inner relations, and any option for the choice of the weight relations. Apel's program is being implemented for optional raw data or moment data input, and for using the revised version (10) of the multiplicative factor system $W_i s_{ij} p_{ij}$.

2. The intermediate position of NIPALS soft models between data analysis and traditional model building

The position at issue is somewhat flexible. We shall consider four version of NIPALS soft modelling, ranging from clearly data-oriented to more "hard" modes of specification.

2.1. *The operative specification* 1.2 *and the estimation* 1.4
The NIPALS approach is distinctly different from data analysis in being based on the notion of model. A NIPALS soft model is conceptually defined by its arrow scheme. Otherwise, the specification 1.2 and estimation procedure 1.4 are data-oriented, inasmuch as the specification formulas (1)–(4) refer to the data, and the ensuing estimation procedure 1.4.3 operates on the data.

Each of the following Sections 2.2–2.5 develops and supplements the soft modelling approach as presented in 1.1–1.4.

2.2. *Population concepts of the model: moments of first and second order.*
Formulating the model in terms of the population, this version takes (2)–(4) to be estimates of the latent variables and the structural relations as defined in the population. For example,

(11a–c) $$x_h = \pi_{1h0}+\pi_{1h}\xi+\nu_{1h}; \quad y_j = \pi_{2j0}+\pi_{2j}\eta+\nu_{2j}; \quad z_k = \pi_{3k0}+\pi_{3k}\zeta+\nu_{3k}$$

is the counterpart in the population to the block structure (3a–c). In the population the indicators are specified by their expectations and their correlation and dispersion matrices, say

(12a–c) $$L^* = [E(x_h), E(y_j), E(z_k)], \quad R^*(x, y, z), \quad V^*(x, y, z).$$

The population parameters of a soft model, such as the theoretical coefficients $\beta_{30}, \beta_{31}, \beta_{32}$ of the inner relation (3d), are obtained by the estimation procedure 1.4.1–1.4.8, using the population moment data (12), and proceeding just as in 1.4.8.

2.3. *Predictor specification of the structural relations*
This version of the model design includes, in addition to 2.2, the specification of the theoretical relations as *predictors*, that is, conditional expectations; [21]. For example, the predictor specification of the population counterpart to the inner relation (3d) reads

(13) $$E(\zeta|\xi, \eta) = \beta_{30}+\beta_{31}\xi+\beta_{32}\eta.$$

(i) A principal aim of NIPALS soft modelling is to obtain between-blocks (inner) relations that are (a) operative for causal-predictive purposes, and (b) relatively stable in various respects. As to (a) the predictor specification (13) makes the inner relations amenable to causal-predictive inference in the same sense as OLS regressions. As to (b), see 3.2.

(ii) On predictor specification of weight relations and inner relations, the estimation procedure 1.4 remains operatively and numerically the same. On predictor specification, under mild additional assumptions, the estimation procedure gives parameter estimates that are consistent in the large-sample sense; [22], [23].

(iii) In econometric model building, structural relations are often referred to as "equations", and their residuals as "errors in equations". Predictor specification is more general, a key difference being that while equation sare (sometimes treated as) reversible, predictor relations (13) are irreversible, for example,

(14) $$E(\xi|\zeta, \eta) \neq (\zeta-\beta_{30}-\beta_{32}\eta)/\beta_{31},$$

except for the special case where the residual ε is vanishing identically.

2.4. *Data simulation*
To simulate data for a soft model as conceptually defined by its arrow scheme, the investigator must specify the parameters of the block structure and the joint probability distribution of the latent variables. To illustrate by Model C331, we shall simulate $N = 100$ cases, assuming that the latent variables have a jointly normal distribution with zero means and specified correlation matrix R^*, subject to independent observations over the 100 cases. First, case values ξ_n, η_n, ζ_n ($n = 1, \ldots, 100$) are generated for the latent variables on the basis of their trivariate normal distribution. Second, case values x_{hn}, y_{jn}, z_{kn} for the indicators are generated from the block structure (11a–c), using the specified parameters, the simulated case values of the latent variables and three sets of simulated residuals, v_{1n}, v_{2n}, v_{3n}.

2.5. *"Hard" modelling*
The data simulation 2.4 has carried us to the realm of traditional model building, with its "hard" assumptions on the distributional properties of the model.

Reference is made to two problem areas where theoretical properties of soft models can be explored by means of simulated data.

(i) *Consistency.* The NIPALS estimates based on a finite sample of size N will, as $N \to \infty$, tend to the estimates based on the moment data of the population, if we disregard exceptional parameter values that are discontinuous in the limit.

(ii) *Consistency at large.* Subject to further assumptions on the residuals v_{1h}, v_{2j}, v_{3k}, and assuming that there are many observables in each block (H, J, K large), and that the sample size N is large relative to the block size, the NIPALS parameter estimates will approximate the teoretical parameters. Special cases: principal components and canonical correlations; cf. [13], [14].

3. NIPALS soft modelling: A beginning of applications

Attempts to break away from the "hard" assumptions of traditional model building are "in the air". Path models with latent variables rapidly came to the fore in sociology in the 1960's, and are now gaining momentum in other social sciences. The NIPALS approach to path models with latent variables thus in its beginnings, both with regard to theory and applications.

3.1. In the wide realm of potential applications of NIPALS soft modelling to interdisciplinary and other complex problems, large areas are *terra incognita* to quantitative analysis. For pioneering work in virgin fields, reference is made to Adelman *et al.* [2], who use Adelman–Morriss's cross section data ([3], [4]) on 74 developing countries to estimate NIPALS three or four-block models for the relationships between natural resources, social factors, political factors, and economic growth. Meissner *et al.* [15] in another pioneering application use NIPALS three to five-block models for the construction of an ecological-economic model for region Hessen.

Noonan *et al.* [17] and Noonan–Wold [18] have applied NIPALS model C331 to the I.E.A. data bank on school pupils' achievements. This is a problem area that for more than two or three decades has been explored by quantitative methods. The NIPALS analysis suggests that previous school survey research has tended to systematically under-rate the influence of the school conditions as compared with the home background.

3.2. The applications of NIPALS soft modelling thus far available are encouraging. From the point of view of subject matter analysis the numerical results are plausible, and give new vistas on the interdisciplinary problems under analysis. In accordance with 2.3 (ii) the numerical results meet the principal aim to obtain between-blocks (inner) relations that are amenable to causal-predictive inference.

As regards matters of statistical technique, the studies in 3.1 are unison in showing that the inner relations and the loadings have a high degree of stability with regard to various tests. As to the weight relations it is shown that the parameters are more stable for A-type than for B-type models. For NIPALS one or two block models the estimation procedure 1.4 converges with probability one; [7]. For larger models, convergence of the procedure has never been a problem in applications to real-world models and data.

4. NIPALS soft modelling: Generalization aspects

The NIPALS approach in Sections 1–2 invites to generalizations in several directions. We shall here refer to extensions where current research has given tangible results.

4.1. *Nonlinearities in the variables*

As is usually the case in least squares modelling, it is direct matter to generalize the design in Sections 1–2 so as to cover nonlinearities in the variables, whereas nonlinearities in the parameters are much more of a problem. For example, transforming the manifest variables by forming squares, inverses, logarithms, etc., the transformed observables may be used as fresh indicators.

4.2. *Hybrids of path models with manifest and latent variables*

If a block of manifest variables in a NIPALS soft model contains just one observable, the latent variable of the block will reduce to this indicator. Hence when given an ordinary path model where all variables are directly observed, we may form hybrid models where one or more of the observables are replaced by latent variables indirectly observed by a block of indicators.

4.3. *Latent variables in two or more dimensions*

To repeat from 1.3.1 (i) the first principal component and the first canonical correlation are special cases of NIPALS soft modelling. Using the residuals of the block

structure as data input, the NIPALS algorithms for these models give principal components and canonical correlations of second order; that is, latent variables in two dimensions, and the procedure can be repeated to give latent variables in three or more dimensions. The device is of general scope in NIPALS soft models. For each block of indicators, whether the weight relations are designed Mode A or B, the consecutive latent variables are mutually uncorrelated. Each new dimension will bring a new term into the block structure. As in canonical correlations, the consecutive dimensions will give different numerical coefficients in the inner relations.

4.4. *NIPALS soft models where the inner relations form an interdependent (ID) system.*

In Classical ID systems the structural relations do not allow predictor specification, and therefore are not amenable to the same causal-predictive interpretation as ordinary regression relations; [21]. Reference is made to REID (Reformulated) and GEID (General ID) systems, and to the estimation of REID–GEID systems by the FP (Fix-Point) method; [16]. REID systems establish predictor specification of the structural relations by replacing explanatory endogenous variables by their predictors; GEID systems generalize classical and REID assumptions about the residual correlations; FP is an iterative partial least squares procedure.

The estimation technique 1.4 is now being adapted so as to cover NIPALS soft models where the inner structural relations are interdependent. The adaptation transforms the system of inner relations from Classical ID to REID-GEID form, and makes combined use of the NIPALS alorithm 1.4 and the FP procedure; [26], [27], [29*].

5. Comparative aspects: NIPALS soft modelling vs. data analysis and Maximum Likelihood methods

5.1. *Data analysis*

Principal components, canonical correlations, and Hauser's model [10] — in NIPALS versions: Models A1, B2, and C2 — are cases of overlapping between NIPALS soft modelling and data analysis. In these simple models numerically the same results can be obtained by different statistical methods. For C331 and more complex NIPALS models there are no counterparts in data analysis, as is only natural since the arrow scheme and the concept of model are here more essential for the purpose of squeezing information from the data.

5.2. *Maximum Likelihood (ML) methods*

Just as the first principal component is the simplest case of NIPALS soft modelling, the first factor in factor analysis is the simplest case of an array of path models with latent variables as developed by K. G. Jöreskog, using ML methods; see the program paper [12], with references to earlier works. The NIPALS soft approach and the ML "hard" approach to path models are similar with regard to the arrow scheme and also in other respects; yet there are distinct differences. To specify:

5.2.1. *Explicit estimation vs. elimination of the latent variables*. In NIPALS soft modelling the latent variable of each block is estimated as a weighted average of the observables in the block. Hence when the weights are estimated, the NIPALS approach gives case values for each latent variable in terms of the case values that the data contains for the observables. The case values can be exploited in various ways for testing and further development of the model.

In ML modelling the latent variables are eliminated in the course of the estimation of the inner and outer structural relations. Hence no case values can be obtained for the latent variables. This I see as a serious disadvantage relative to NIPALS soft modelling.

5.2.2. *Identification*. Thanks to the explicit estimation of the latent variables, NIPALS soft models are always identifiable. In corresponding ML models, on the other hand, the elimination of the latent variables during the estimation is a transformation of the model, and as always the transformation involves some loss of information with regard to the identification of the model. Again to specify:

(i) The ML versions of Models C2, C331 and more complex NIPALS soft models are not identifiable, unless ancillary assumptions are incorporated that are not implied in the arrow scheme; [9]. The more complex the model, the more ancillary assumptions are needed.

(ii) As noted in 1.4.5 it makes no problem in NIPALS soft modelling to include the location constants of variables and relations in the analysis. In corresponding ML models the location parameters add considerably to the intricacy of the identification problems; see [19].

5.2.3. *Hypothesis testing*. Speaking broadly, the ML methods of hypothesis testing are distribution oriented, providing significance levels for likelihood-ratio and other tests on the model; see [12]. NIPALS soft modelling, in contrast, uses data oriented methods. Specific reference is made to perturbations analysis [1], and to cross validation, [28]. Having been implemented for principal component models, this last-mentioned method allows straight forward generalization to multi-block NIPALS soft models.

5.3. *The scope of NIPALS vs. ML approaches to path models with latent variables*. The two approaches are complementary rather than competitive. The "hard" ML approach is appropriate in microanalysis and relatively simple models, where the observational experience provides prior knowledge about the distributional assumptions as well as the assumptions needed for identifiability (see 5.2.2). The "soft" NIPALS approach is primarily intended for macroanalysis of nonexperimental data on interdisciplinary and other complex problems. Between these very different situations there is a wide realm of intermediate problem areas where the reach and limitation of NIPALS "soft" and ML "hard" modelling is largely unexplored, and where in due course the success or failure of applications to real-world problems and data will be decisive for the verdict.

Acknowledgements. This paper is a write-up of four lectures at the Stefan Banach International Center of Mathematics, Warsaw, in October 1976, and draws also from the planning of a course on NIPALS soft modelling at the University of Pennsylvania in the Spring Term 1977. The paper is expository. What is new lies mainly in the arrangement of the material, the spelling out in 1.3 of four principles of NIPALS modelling, and in 2.1–2.4 of four intermediate positions of NIPALS modelling between data analysis and traditional model building. The sign rule (10) removes an inadequacy in previous versions of the double factor system.

The research project on NIPALS soft modelling under which this paper belongs has received economic support from The Tercentenary Fund of the Bank of Sweden, the Scientific Council of the Central Bureau of Statistics, Stockholm, and recently from the Stiftung Volkswagenwerk, Hannover.

* The proofs were received in March 1980, more than three years after the manuscript was submitted to the editors. The developments of soft modelling in this period have been taken into account in the proof-reading by changing obsolete terminology and notations, by some clarifying reediting, by adding references [29*]–[30*], but without making substantive changes.

References

[1] I. Adelman and Å. Abrahamsson, B. Areskoug, *Perturbation analysis of NIPALS parameter estimates*, Ch. 6 in [25], 1975.

[2] I. Adelman, and Å. Abrahamsson, B. Areskoug, L.-O. Lorentzson, J. Wallmyr, *Applications of method I–II* [*Clustering and modelling using the NIPALS approach*] *to Adelman–Morriss' data*, Ch. 4 in [25], 1975.

[3] I. Adelman and C. T. Morriss, *Society, politics, and economic development: A quantitative approach*, Johns Hopkins Press, Baltimore 1967; revised ed. 1971.

[4] —, —, *Economic growth and social equity in developing countries*, University Press, Stanford 1973.

[5] H. Apel, *Simulation sozio-ökonomischer Zusammenhänge-Kritik und Modifikation von Systems Dynamics*, doc. disser., J. W. Goethe Univer., Frankfurt/Main, 1977.

[6] B. Areskoug, *Computer program MIMIC*, Dep. of Statistics, Univer. of Göteborg, 1974/75.

[7] B. Areskoug and E. Lyttkens, H. Wold, *Six models with two blocks of observables for one or two latent variables*, Research Rep. 1975: 3, Dep. of Statistics. Univer. of Göteborg 1975.

[8] G. Bray, *On a new interpretation of the models of H. Wold with latent variables*, Report 1976: 1, Faculty of Economic Sciences, Univer. of Rennes, 1976.

[9] J. Dupačová and H. Wold, *NIPALS vs. maximum likelihood modelling*, Ch. 8 in [25].

[10] R. M. Hauser, *Family, school, neighborhood factors in educational performances in a metropolitan school system*, doc. disser., Univer. of Michigan, 1968.

[11] B. Hui and R. Hausman, *Computer program KKIHUI*, Dep. of Statistics, Wharton School, Univer. of Pennsylvania.

[12] K. G. Jöreskog, *Causal models in the social sciences: the need for methodological research*, pages 47–68 in *Uppsala University 500 years*, 7, Acta Universitatis Upsaliensis, 1976.
[13] E. Lyttkens, *On the fix-point property of Wold's iterative estimation method for principal components*, pages 335–350 in *Multivariate analysis*, edit. P. R. Krishnaiah, Academic Press. New York 1966.
[14] —, *Regression aspects of canonical correlation*, J. Multivariate analysis 2 (1972), pp. 418–439.
[15] W. Meissner and H. Apel, W. Fassing, M. Tschirschwitz, *Ökonomische Aspekte des Umweltproblems*, Dep. of Economics, J. W. Goethe Univ., Frankfurt/Main.
[16] E. J. Mosbaek and H. Wold, with contributions by E. Lyttkens, A. Ågren, L. Bodin, *Interdependent systems. Structure and estimation*, North-Holland Publ., Amsterdam 1970.
[17] R. Noonan, R., and Å. Abrahamsson, B. Areskoug, L.-O. Lorentzson, J. Wallmyr, *Applications of methods I–II* [*Clustering and modelling using the NIPALS approach*] *to the I.E.A. Data Bank*, ch. 5 in [25], 1975.
[18] R. Noonan and H. Wold, *NIPALS path modelling with latent variables. Analysing school survey data using nonlinear iterative partial least squares*, Scandinavian J. of Educational Research 21 (1977), pp. 33–61.
[19] D. Sörbom, *Statistical methodology for model building with latent variables*, doc. disser., University of Uppsala 1976.
[20] P. Whittle, Written communication, 1977.
[21] H. Wold, *Ends and means in econometric model building. Basic considerations reviewed*, pages 355–434 in: *Probability and statistics. The Harald Cramér Volume*, edit. U. Grenander, Almquist & Wiksell, Stockholm 1959; Wiley, New York 1960.
[22] —, *On the consistency of least squares regression*, Sankhya A25, Part 2, pp. 211–215, 1963,
[23] —, *Toward a verdict on macroeconomic simultaneous equations*, Pontifical Academy of Sciences. Vatican City, Scripta Varia 28 (1965), pp. 115–166.
[24] —, *Nonlinear estimation by iterative least squares procedures*, pages 411–444 in: *Research papers in statistics, Festschrift for J. Neyman*, edit. F. N. David, Wiley, New York 1966.
[25] — (edit.), *Modelling in complex situations with soft information. The NIPALS (Nonlinear Iterative Partial Least Squares) approach*, Third World Congress of Econometric Society, Toronto, Canada, 21–26 August 1975. Also: Research Rep. 1975:5, Dep. of Statistics. University of Göteborg, 1975.
[26] —, *On the transition from pattern cognition to model building*, Part I, in: *Festschrift Oskar Morgenstern*, edit. R. Henn & O. Moeschlin, 1977.
[27] — *Open path models with latent variables*, in: *Festschrift Wilhelm Krelle*, edit. H. Albach, E. Helmstedter & R. Henn; Mohr, Tübingen, Springer, Berlin 1977.
[28] S. Wold, *Pattern recognition by means of disjoint principal components models*, Pattern Recognition, 8 (1976), pp. 127–139.
[29*] B. S. Hui, *The Partial Least Squares approach to path models of indirectly observed variables with multiple indicators*, doc. disser., University of Pennsylvania, 1978.
[30*] H. Wold, *Model construction and evaluation when theoretical knowledge is scarce. An example of the use of Partial Least Squares*, in: *Evaluation of Econometric Models*, edit. J. Kmenta, J. B. Ramsey, Academic Press, New York (in press). Prepublication version: Cahier 1979.06, Department of Econometrics, University of Geneva.

Presented to the semester
MATHEMATICAL STATISTICS
September 15–December 18, 1976

icm©

MATHEMATICAL STATISTICS
BANACH CENTER PUBLICATIONS, VOLUME 6
PWN—POLISH SCIENTIFIC PUBLISHERS
WARSAW 1980

SOME REMARKS ON LARGE DEVIATIONS FOR WEIGHTED SUMS IF CRAMÉR'S CONDITION IS NOT SATISFIED

W. WOLF

Technical University of Dresden, G.D.R.

1. Introduction

1.1. We consider a sequence of independent identically distributed random variables $X_1, X_2, \ldots$ with $EX_1 = 0$ and $D^2X_1 = 1$ and a double array $\{a\} = \{a_{nk}, 1 \leqslant k \leqslant n, 1 \leqslant n < \infty\}$ of nonnegative numbers. We want to study the asymptotic behaviour of the probabilities

$$P\{a_{n1}X_1 + \ldots + a_{nn}X_n > x\} \quad \text{or} \quad P\{a_{n1}X_1 + \ldots + a_{nn}X_n < -x\} \tag{1.1}$$

in the case where if $n \to \infty$ also $x = x(n) \to \infty$. Large deviation theorems for weighted sums under Cramér's condition were studied by S. A. Book [1], [2], L. Saulis and V. Statulevičius [6]. Our aim is to derive asymptotic representations for the probabilities (1.1) if Cramér's condition is not satisfied.

1.2. In the following, g always denotes a function with the following properties: $g(x)$ is nondecreasing and continuous if $x > C(g)$ and satisfies the conditions

$$\varrho(x)\ln x \leqslant g(x) \leqslant C^*(g)x^{\alpha}, \quad 0 < \alpha < 1 \tag{1.2}$$

and

$$g(x)x^{-1} \text{ is strictly decreasing.} \tag{1.3}$$

(Here $\varrho(x)$ is an arbitrary monotone increasing function with $\lim_{x\to\infty} \varrho(x) = \infty$, $C(g)$ and $C^*(g)$ are positive constants depending on g.)

Furthermore, let the array $\{a\}$ satisfy the following condition (see [6]):

There exist numbers δ and β, $0 < \delta \leqslant 1$, $0 < \beta \leqslant 1$, such that, for every sufficiently large n, for at least δn of the a_{nk}'s the inequalities

$$a_{nk} \geqslant \beta\gamma_n \tag{1.4}$$

hold; here

$$\gamma_n = \max_k\{a_{nk},\ 1 \leqslant k \leqslant n\}. \tag{1.5}$$