

This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Evaluation of Econometric Models

Volume Author/Editor: Jan Kmenta and James B. Ramsey, eds.

Volume Publisher: Academic Press

Volume ISBN: 978-0-12-416550-2

Volume URL: <http://www.nber.org/books/kmen80-1>

Publication Date: 1980

Chapter Title: Data Analysis by Partial Least Squares

Chapter Author: Fred L. Bookstein

Chapter URL: <http://www.nber.org/chapters/c11694>

Chapter pages in book: (p. 75 - 90)

# Data Analysis by Partial Least Squares

FRED L. BOOKSTEIN

CENTER FOR HUMAN GROWTH AND DEVELOPMENT  
AND INSTITUTE FOR SOCIAL RESEARCH  
UNIVERSITY OF MICHIGAN  
ANN ARBOR, MICHIGAN

## 1. Introduction. Latent Variables and Soft Models

For the phenomena social scientists study, causation generally operates at the level of events: decisions, outcomes, and opportunities. But were each event measured separately, there would be no such activity as prediction, only prophecy. In accounting for patterns of events in any predictive way, causal theories must depend on measurable attributes—of individuals, institutions, interactions—presumed stable over events. Then the variables we measure are usually misspecified for the causal schemes we believe to govern the process of their interrelations; they are *all* proxies, all at the wrong level of aggregation.

In modern quantitative practice the method of latent variables emerges as a general response to this perplexity. By way of compensating for the misspecification of any causally relevant empirical attribute, we measure it variously and repeatedly. Each “variable” becomes, in practice, a *block* of many *items*. A *latent variable* (Lv) is a scale score which combines the items of a block into a single quantity for arithmetic use later in a particular causal model. The Lv is formed by three considerations: the items whose causal force it embodies; the other variables, observed or latent, in the causal scheme; and the details of the algorithm by which we generate the scale scores to combine the items of a block in a single aggregate. Latent variables may be linear expressions in the items of a block, integers, switches, or any other sort of statistical artifact required by a particular style of prediction.

Over several years of exposition and collaboration Herman Wold has developed for latent variable analyses a set of conventions which he calls

*soft modeling*. By “soft” here Wold means weak in assumptions, undemanding. His soft models presume several blocks, each an assemblage of items that are all proxies of one Lv. The blocks are related by an *arrow diagram*, a causal chain specifying that certain Lvs must have an expected value depending linearly on the values of others. Each Lv is to be an explicit linear combination of the items of its block. No other information is supplied, no other assumptions are made.

To estimate such a model is to fix coefficients of two sorts: the regression weights (“inner relations”), whereby values of an Lv “affect” values of others further down the causal chain, and the item weights (“outer relations”), which describe the manner in which the items of a block severally determine the Lv which represents them. Whereas conventional multiequation estimators require a stringent parametric model, namely, a family of joint distributions for all the items in all the blocks, the soft specification is translated into a collection of partial linear models separately, almost trivially, analyzed by ordinary least squares regressions. This unexpected tactic is directed by two formal themes, suboptimality and simplicity.

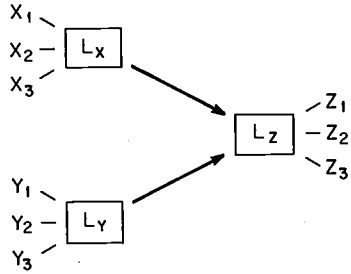
1. *Suboptimality* refers to the formal disaggregation of the computations. Each block in the model, its items together with its eventual Lv, can be interpreted as the basis of a *submodel* consisting just of that block and the other blocks with which it communicates, that is, to which it is linked by explicit arrows of the causal scheme. In Wold’s prescriptions, each Lv is related only to the other Lvs of its submodel together with the items of its own block. Nevertheless, since the submodels overlap, the result is an interdependent system of equations between latent variables.

2. For the sake of *simplicity* any Lv is characterized in terms of projections of the Lvs of its submodel upon the items of its own block. The minimum-distance property of projection onto a block is the *only* optimization principle invoked in the course of soft modeling.

The Lvs of a soft model may be jointly characterized using a complex, nonlinear operator for which the vector of all estimated item weights (outer relations) serves as a *fixed-point*. Soft estimation, then, does not resemble at all the search for zeroes of certain derivatives which characterizes the estimation of “harder” models. In its stead, in order to find the fixed-point of a soft model, Wold proposes a succession of regressions and linear combinations within submodels—a cycle of replacement of “old” Lvs (earlier estimates) by new—which seems always to converge.

In this essay I will explore the role of these two formal themes in Wold’s exposition. Regarding his prescriptions for submodel regressions, I will indicate which are dictated by the needs of suboptimality or simplicity and which are left for determination by content of the theory, model, or data at hand. The procedures I derive from the principles are somewhat more flexible than Wold’s own.

Fig. 1. Typical arrow diagram. A latent variable  $L_X$ , proxied by indicators  $X_1, X_2, X_3$ , and an Lv  $L_Y$  proxied by  $Y_1, Y_2, Y_3$ , jointly and linearly determine the expected value of an Lv  $L_Z$  proxied by items  $Z_1, Z_2, Z_3$ .

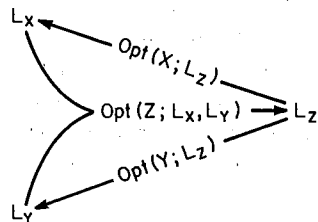


2. The Command Diagram

We begin with an arrow diagram. The specimen in Fig. 1 sets out three blocks of items with corresponding Lvs. The Lvs are to capture X-ness, Y-ness, and Z-ness in the context of a joint linear determination of Z-ness by X-ness and Y-ness. (Latent variables here are denoted by the letter  $L$  subscripted by the block name, e.g.,  $L_X, L_Y, L_Z$ .)

Under Wold's rules the arrow diagram is transformed into a series of operations replacing each tentative Lv by a new linear combination of the items in its block. The new Lv is not a function of the old Lv at all, but is rather a function of the Lvs of the other blocks of its submodel as they relate to the items of its own block. The new Lv is always some sort of orthogonal projection; we may disregard details by using the general operator  $Opt$ , without specifying its precise functional form. The operator  $Opt$  has two arguments. The first is the block for which the operator's output is the new Lv; the second is a list of all Lvs linked to the output block by direct paths in the arrow diagram—a roster of the submodel. In this manner the arrow diagram in Fig. 1 is replaced by the command diagram in Fig. 2, which makes explicit the arguments and outputs of the various operations, that is, the flow of partial regressions over the diagram. The various  $Opt$  commands have been written out over paths connecting their arguments to their outputs, the Lvs for the next round of iteration. For example, the path  $Opt(X; L_Z)$  is a regression, of form not yet specified, which computes a new Lv  $L_X$  by relating the current Lv  $L_Z$ , a linear combination of items  $Z_1 \dots Z_3$ , to the

Fig. 2. Command diagram transcribing Fig. 1. Each operation is an optimization for a new latent variable as a function of all the Lv's communicating with it. The arrows express association of inputs and outputs, not causation.



items of the  $X$ -block. The direction of the arrows in the command diagram expresses the order of operations and outputs, not the flow of causation, which will be embodied in the detailed formulation of the functions Opt.

The command diagram directly implies the algorithm to be used for estimating both sorts of coefficients, regression weights and item coefficients, as follows.

1. Begin by defining each latent variable as an arbitrary linear combination of the items in its block. Reasonable starting values might be  $L_x = \sum X_i$ , etc.

2. Determine the exact form of each Opt command in terms of some short sequence of regressions and linear combinations, utilizing the considerations set forth later in this exposition.

3. Execute all the commands of step 2 just once, without updating the Lv formulas.

4. Have we arrived at the fixed-point?—are the Lvs consistent with their joint characterization to some preset tolerance? If so, we are done; otherwise,

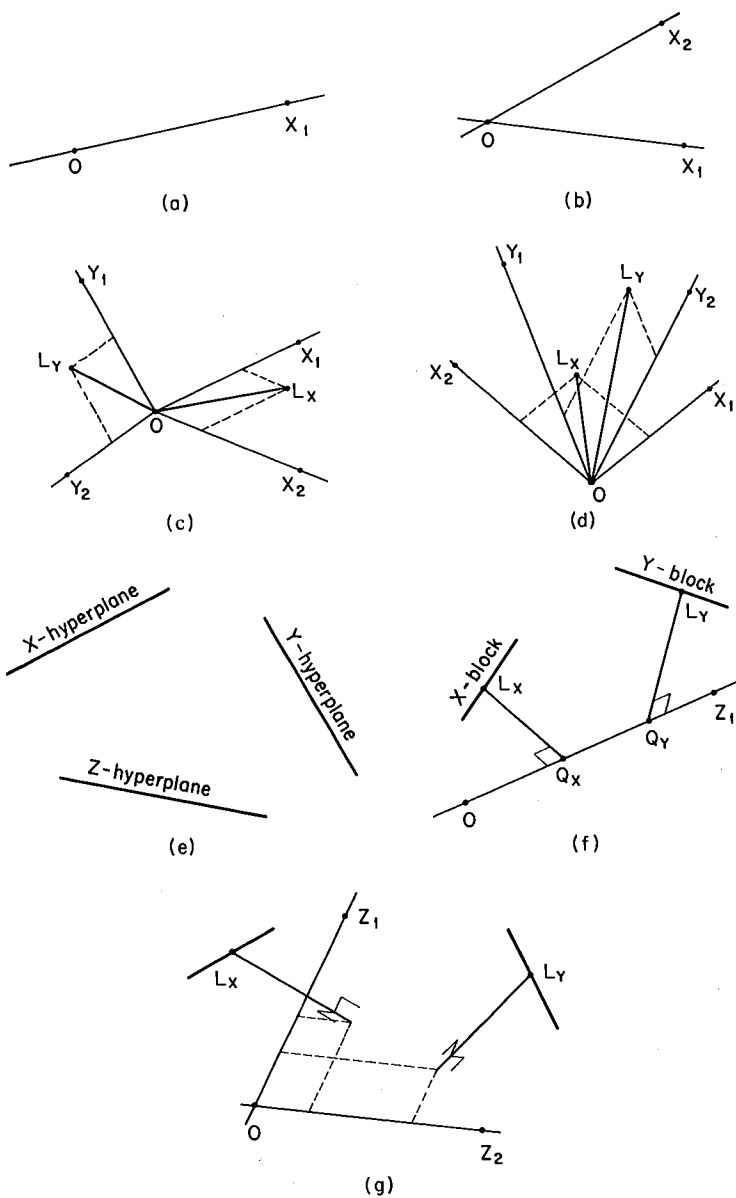
5. Replace each Lv by the output of the appropriate Opt operator, scaled to unit variance, and return to step 3.

In practice these algorithms always converge. At convergence, each Lv is a projection onto its block of the Lvs from the other blocks of its sub-model. Only in these projections does optimality lie; otherwise the estimates are characterized by their self-consistency as tested in step 4.

### 3. Polygon Diagrams

The intrinsic geometric content of soft techniques stems from the equivalence of regression and orthogonal projections from point to hyperplane. The fixed-point of a soft model can be expressed in terms of the mutual orientations among hyperplanes in a high-dimensional space. Their soft modeling is mainly geometry, so we should be able to visualize what we are doing.

I shall try to demonstrate how to distinguish between alternative forms of the Opt operator with diagrams on flat paper, but to do so I shall need to establish some visual conventions. Figures 3a–g illustrate the standard constructions to which I shall be referring. Until one notices the apparent inconsistency of the right-angle ( $\Gamma$ ) symbols scattered throughout the diagrams, the figures appear to be of points, lines, and planes from ordinary three-dimensional space. But, in fact, all of the objects depicted lie in the



**Fig. 3.** Elements of the polygon diagram. (a) The line of multiples of an item  $X_1$ . 0 is the zero vector. (b) The hyperplane corresponding to the X-block, viewed plane-on. (c) Two hyperplanes, each with a latent variable, viewed from in between. For any vector of a block the dotted lines connect it to a basis for its block. (d) The same, viewed through the X-hyperplane to the Y-hyperplane. (e) Three hyperplanes all edge-on. (f) Two projections from hyperplanes onto a line. Orthogonality of projection onto a rank-1 subspace is shown by the symbol  $\perp$  drawn in perspective. (g) Two projections from hyperplanes onto a hyperplane viewed plane-on. Orthogonality of projection onto a higher-rank subspace is shown by a combination of symbols  $\perp$  for each basis vector separately.

space dual to that of the items in an analysis, namely the vector space of their linear combinations. Points in these diagrams are particular scores, such as the Lvs themselves; lines and planes depict subspaces spanned by the items of a block or by sets of Lvs. The diagrams are drawn in no particular coordinate system at all, though the representation of regression as explicit orthogonal projection suggests a set of orthonormal axes as the appropriate basis. In terms of the items themselves, the variables of the raw data matrix, the inner product here is defined not in terms of the Euclidean cosine, but in terms of the familiar covariance matrix  $\Sigma$ . Submatrices  $\Sigma_{XX}$ ,  $\Sigma_{XY}$ , etc. of  $\Sigma$  incorporate the covariances within and between the separate blocks of the model.

Let us proceed through the frames of Fig. 3.

Figure 3a shows a typical line of our space. Here it is the set of all multiples of an item,  $X_1$ . Note the zero vector 0, which is represented by a point on this and all other lines and higher-rank subspaces.

Figure 3b shows a hyperplane of this space, viewed plane-on so that we can see two of the items making it up. Note that these axes are oblique, as items  $X_1$  and  $X_2$  are, in general, correlated. Any block of the soft model can be drawn as a hyperplane whose points are all the possible Lvs for that block. The algorithms of soft modeling determine one Lv per block—one point per hyperplane which is in the proper relation to the Lvs of the other blocks of its submodel.

Figure 3c shows two hyperplanes, after the fashion of Fig. 3b, with an Lv in each one. We indicate the (hyper)plane in which a point resides by dropping dashed lines to the lines determined by the items spanning that (hyper)plane. Note that these planes can intersect only at the point 0; elsewhere they avoid each other like two lines in a plane. We are looking at them "from in between."

Figure 3d shows the same two hyperplanes, from a viewing position now "through the  $X$ -block to the  $Y$ -block." Certain constructions will appear more legible from this perspective.

Figure 3e shows three hyperplanes, all edge-on to the viewing eye, in some inconceivable high-dimensional rotation. These three entities intersect in a single point only, the origin of coordinates, and do not intersect, even in pairs, anywhere else at all. One must imagine them curving around to the point of concurrence through dimensions unrelated to flat paper or even physical space. It should be clear why the point 0 is not drawn; there is no logical place to put it.

Figure 3f shows two regressions, from hyperplanes representing two blocks of a model, onto the line of multiples of an item  $Z_1$  from the  $Z$ -block. An Lv  $L_X$ , residing within the hyperplane corresponding to the  $X$ -block of

the model (here seen edge-on), is regressed on a single item  $Z_1$ , that is to say, projected orthogonally onto the line from 0 through  $Z_1$ . (Recall that 0 is included in both blocks as well as in the line.) The predicted value of  $L_X$  is the vector  $Q_X$ , some multiple of  $Z_1$ ; the residual of that regression is the vector difference  $L_X - Q_X$ , orthogonal to the lines of multiples of  $Z_1$ , as shown. Note that the symbol for orthogonality of two lines is drawn in perspective.

Similarly, in this figure the latent variable  $L_Y$  is regressed onto the same single item  $Z_1$ . The perception that these projections come from different directions is to be encouraged; that they seem to come from different distances is tempered by the difficulty of representing all Lvs at the same distance from 0 (as they are all normalized to the same variance).

Figure 3g shows two regressions onto a hyperplane viewed plane-on. The symbol for orthogonality of a projection onto a hyperplane is a combination of  $\Gamma$ 's, in perspective, indicating orthogonality to the basis vectors of the plane of view. The "normals" for these two projections (residuals of the regressions) come from different directions in hyperspace, resulting in a paradoxical perspective for the diagram as a whole.

A soft algorithm terminates when a loop of Opt operations returns to the very latent variables, hyperplane by hyperplane, with which it began. The fixed-point itself, though really a vector, can then be drawn as a whole *closed polygon* in the diagram, a path whose edges bear Lvs at one end and right angles at the other, as in Fig. 9b and 10c. In terms of its symbols of perpendicularity the polygon diagram indicates the precise form of each Opt command to be programmed for calculation through regression analysis. For instance, the model of Fig. 9a-b is executed by the following command sequences:

0. Set  $L_Z = \sum Z_i$ , normalized to variance 1.
1. Compute  $L_X$  by determining the predicted value of  $L_Z$  given the X-items.
2. Compute  $L_Y$  by determining the predicted value of  $L_Z$  given the Y-items.
3. Compute  $L_W$  by determining the predicted value of  $L_Z$  given  $L_X$  and  $L_Y$ .
4. Compute  $L_Z^*$  by determining the predicted value of  $L_W$  given the Z-items.
5. Set  $L'_Z = L_Z^*/\sigma_{L_Z}$  where the denominator is the observed variance of  $L_Z^*$ .
6. TEST. If  $L'_Z$  is sufficiently close to  $L_Z$ , exit; otherwise,
7. Set  $L_Z = L'_Z$  and go to step 1 again.



The flow of computation is drawn in Fig. 9c by associating the step numbers from the preceding list to the appropriate edges in the polygon diagram, paths from dependent variables to predicted value which link the hyperplanes of the model.

#### 4. Precise Forms of the Operator “Opt” for the Relationship between Two Latent Variables

Using the elements introduced in Fig. 3 we can explore a diversity of Opt commands by combining projections in various ways.

Consider first the simplest submodels, those containing only two blocks of items. These correspond to Opts whose second argument is a single Lv only. In the absence of formal asymmetries among the items of the output block, there seem to be only two reasonable ways to proceed, shown in Fig. 4 for the case of the command  $L_X = \text{Opt}(X; L_Z)$  from Fig. 2. As the notation indicates, either construction is intended to provide a tentative Lv in the X-hyperplane. That is, a linear combination of the X-items is to be determined with respect to a tentative Lv  $L_Z$  in the Z-hyperplane.

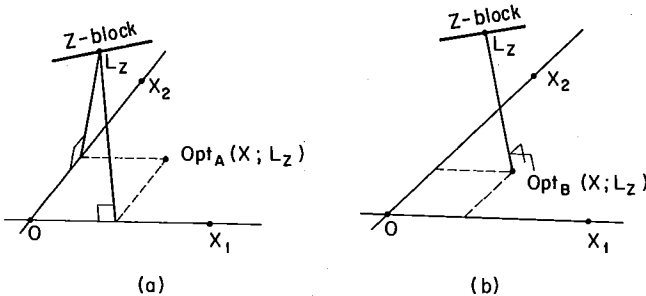


Fig. 4. The two modes of the Opt command for a block  $X$  communicating with only one other block  $Z$ . (a) Factor estimation mode.  $L_X$  is the sum of the simple regression functions of  $L_Z$  upon the  $X$ -items separately. (b) Linear estimation mode, orthogonal projection of  $L_Z$  upon the  $X$ -hyperplane.

Mode A projects  $L_Z$  upon the items  $X_i$  of the  $X$ -block separately, then constructs the sum of these projections. When all items are of unit variance, as will be assumed throughout the remainder of this exposition, the result is  $L_X = \Sigma_{XZ}L_Z$ . The geometry of this  $\text{Opt}_A(X; L_Z)$  is diagrammed in Fig. 4a.

Mode B projects  $L_Z$  onto the  $X$ -hyperplane directly, in one operation. The result, shown in Fig. 4b, is  $\text{Opt}_B(X; L_Z) = \Sigma_{XX}^{-1} \Sigma_{XZ}L_Z$ , which differs from  $\text{Opt}_A$  whenever the  $X$ s show any intercorrelation, i.e., when  $\Sigma_{XX}$  is nondiagonal.

To each of these modes, A and B, corresponds a familiar statistical model.  $\text{Opt}_B$  is the multiple regression of  $L_Z$  on the  $X$ -items, that linear combination  $L_X$  which minimizes the root-mean-square of  $\varepsilon$  in the regression equation  $L_Z = L_X + \varepsilon$ .  $\text{Opt}_A$  is factor estimation. If each indicator  $X_i$  is representable by  $a_i L_Z + \varepsilon_i$ , where the  $\varepsilon_i$  are mutually independent random deviates with mean zero, then  $\text{Opt}_A(X; L_Z)$  is the estimator of  $L_Z$  with the smallest summed mean squared  $\varepsilon_i$ .

Though Wold requires us to choose one mode or the other, A or B, for each block, there seems to be no way to decide between them on the basis of observation of distributions or to estimate their relative import for an estimate of the "true situation" by any simple technique. This ambiguity is unfortunate, since the factor  $\Sigma_{XX}^{-1}$  which distinguishes the modes can have a great effect—as the indicators of any block are almost certainly correlated, this inverse will have some large eigenvalues with associated eigenvectors not necessarily uncorrelated with  $\Sigma_{XZ}L_Z$ . To guard against ill-conditioning of  $\Sigma$ , one might, for instance, "deflate"  $\text{Opt}_B$  by an additional multiplication by  $\Sigma_{XX,Z}$ , suppressing dimensions of the  $X$ -hyperplane according to their alienation from the  $Z$ -items. Perplexities of the single-block Opt operation are a prime site for further research into soft modeling.

### 5. Precise Forms of the Operator "Opt" for the Relationships among Several Latent Variables: Approaches to Eigenanalysis

In the previous section we considered whether or not to use the covariance structure  $\Sigma_{XX}$  of the indicators of a single block. When we attend to larger submodels, three blocks or more, there is an analogous choice to be made: we must examine the consequences of ignoring or recognizing the covariances among the connected Lvs themselves.

Wold favours the most straightforward of approaches, simply disregarding the covariances of Lvs. Mode C will combine this strategy with mode A within blocks,<sup>1</sup> mode D with mode B within blocks. Figures 5a and b show the constructions which result for the operation  $L_Z = \text{Opt}(Z; L_X, L_Y)$  from Fig. 2. In either case the output  $L_Z$  is the simple mean of the appropriate  $\text{Opt}_A$ s or  $\text{Opt}_B$ s computed in block  $Z$  with respect to the communicating Lvs separately. By linearity of projection, the Lv  $\text{Opt}_C$  predicts the  $Z$ s from  $L_X + L_Y$ —minimizes mean squared prediction error over all the  $Z$ s separately—while  $\text{Opt}_D$  finds a Lv  $L_Z$  to predict  $L_X + L_Y$  with minimal mean-squared error in the units of  $L_X + L_Y$ .

<sup>1</sup>This is not the mode C of Wold's exposition.

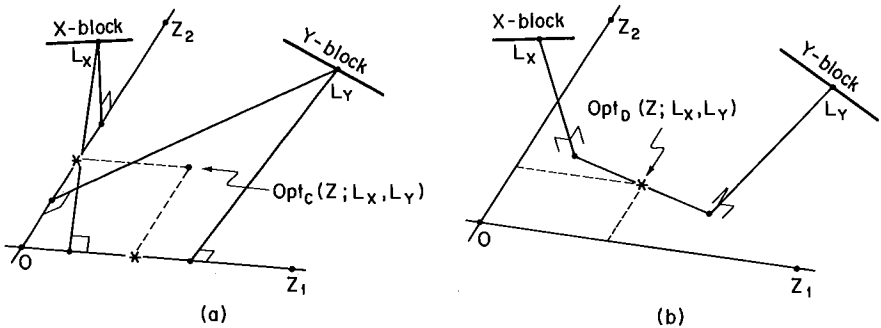


Fig. 5. Opt operators that ignore the covariance structure of the Lvs within a submodel. (a) Factor estimation mode  $Opt_C$ , mean of the separate  $Opt_A$ s for each communicating Lv; sum of the means of all the simple regression functions item by item. (b) Linear estimation mode  $Opt_D$ , mean of the separate  $Opt_B$ s block by block.

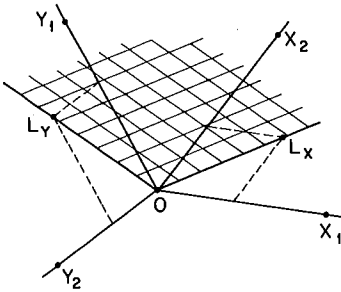
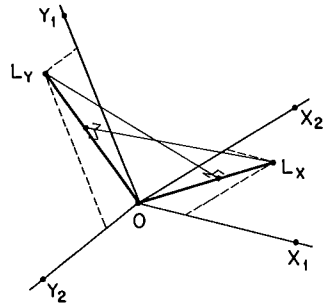


Fig. 6. The plane  $W$  spanned by the Lvs  $L_X, L_Y$  from two separate blocks  $X$  and  $Y$ .

Simple contradiction of this direct route suggests other possibilities in which the communicating Lvs are considered jointly. Mode E will maximize the sum of the squared *simple* correlations of  $L_Z$  with  $L_X$  and  $L_Y$ ; mode F will maximize the squared *multiple* correlation of  $L_Z$  with  $L_X$  and  $L_Y$ . These new optima emerge from computations involving a new hybrid block  $W$  spanned by  $L_X$  and  $L_Y$  themselves.  $W$  is a plane, not a hyperplane—the collection of all linear combinations of  $L_X$  and  $L_Y$  only. It does not lie wholly in either the  $X$ - or the  $Y$ -hyperplane, but rather straddles the “space” between them, as in Fig. 6.

In terms of the matrices  $\Sigma_{WW}$  and  $\Sigma_{WZ}$  of correlations among the Lvs of the  $W$ -plane and between them and the indicators of the  $Z$ -block, the optima of modes E and F are expressible as eigenvectors. The vector  $L_Z$  which maximizes  $R^2(L_Z | L_X, L_Y)$  is the first canonical variable of the  $Z$ -block with respect to the  $W$ -block, the dominant eigenvector of the matrix  $\Sigma_{ZW}\Sigma_{WW}^{-1}\Sigma_{WZ}$  with respect to  $\Sigma_{ZZ}$ ; the vector  $L_Z$  which maximizes  $R^2(L_Z | L_X) + R^2(L_Z | L_Y)$  is easily shown to be a closely related quantity, the dominant eigenvector of the matrix  $\Sigma_{ZW}\Sigma_{WZ}$  with respect to  $\Sigma_{ZZ}$ .

Fig. 7. The polygon diagram for classic two-block canonical analysis. Vector  $L_X$  is in the  $(X_1, X_2)$ -plane, vector  $L_Y$  in the  $(Y_1, Y_2)$ -plane. The expected value of  $L_X$ , when regressed on the items of the Y-block, is proportional to  $L_Y$ ; and the expected value of  $L_Y$  regressed on the X-block is proportional to  $L_X$ .



Now, as eigenvector extractions, these would seem to be excluded from the repertoire of soft modeling. But, in addition to regression, we have another privilege, the passage to the limit, which contributes the needed extension to this class of algorithms. Figure 7 shows, for instance, a two-block model, estimated according to mode B, at convergence. The normalizations after each round of optimization are shown explicitly. In view of the algebraic characterization of mode-B optimization, we may write  $L_X = \lambda \Sigma_{XX}^{-1} \Sigma_{XY} L_Y$ ,  $L_Y = \mu \Sigma_{YY}^{-1} \Sigma_{YX} L_X$ , where  $\lambda$  and  $\mu$  are normalizing factors. Substituting either equation in the other, we see that each of  $L_X$  and  $L_Y$  is an eigenvalue of the appropriate matrix for the problem of canonical variates of the X-block with respect to the Y-block; and although they are eigenvectors, we have computed them by iteration of regressions only.

At convergence, then, the mode-F  $L_Z$  drawn out in Fig. 8b will have maximized  $R^2(L_Z | L_X, L_Y)$ . Note that we cannot write this as  $\text{Opt}_F$  on the

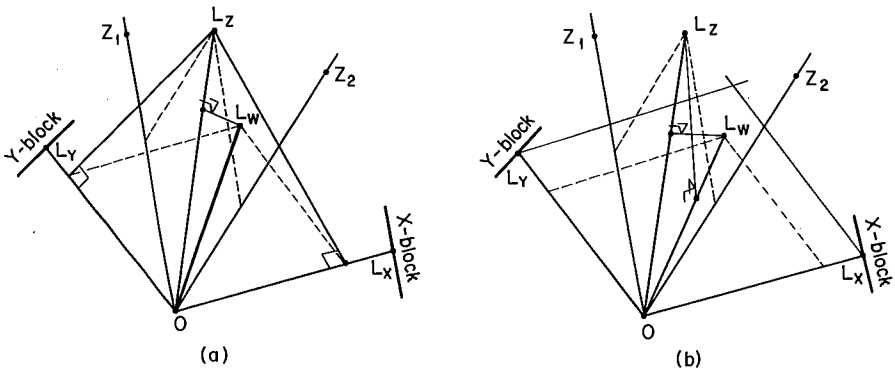


Fig. 8. Polygon diagrams for optimization modes E, F. Vector  $L_Z$  is in the  $(Z_1, Z_2)$ -plane,  $L_X$  in the X-block,  $L_Y$  in the Y-block. Vector  $L_W$  is in the  $(L_X, L_Y)$ -plane. At convergence,  $L_Z$  is the predicted value of  $L_W$  given the items of the Z-block. (a) Computation of  $L_W$  by optimization mode A. The eigenanalysis shows the maximization of the mean-square of the simple correlations of  $L_X, L_Y$  with  $L_Z$ . (b) Computation of  $L_W$  by optimization mode B. The eigenanalysis is that of canonical correlation analysis, showing maximization of the multiple correlation of  $L_Z$  with  $L_X$  and  $L_Y$  jointly.

command diagram, for it is not a finite sequence of regressions—it emerges only at convergence from the succession  $L_W = \text{Opt}_B, L_Z = \text{Opt}_B$ .

Similarly, when the estimation of a model according to the polygon in Fig. 8a has converged, it is to a variable  $L_Z$  which is an eigenvector maximizing  $R^2(L_Z|L_X) + R^2(L_Z|L_Y)$ . This mode-E optimization emerges at convergence from the sequence  $L_W = \text{Opt}_A, L_Z = \text{Opt}_B$ .

The inclusion of the ancillary plane  $W$  in the polygon diagram (or the ancillary block  $W$  in the command diagram) makes it possible to reflect in the operator  $\text{Opt}$  the differences in position of blocks within the causal structure of the model. When determination is joint, as in Fig. 1, optimization should proceed according to mode F (command diagram, Fig. 9a, and polygon diagram, Fig. 9b), in which the maximand for computation of  $L_Z$  is the *multiple* correlation with  $L_X$  and  $L_Y$ . Corresponding to the postulation of a single linear dependency in this submodel, there is a single scalar to be maximized. In a submodel like that of Fig. 10a, where two separate dependencies are being described, I believe optimization should proceed

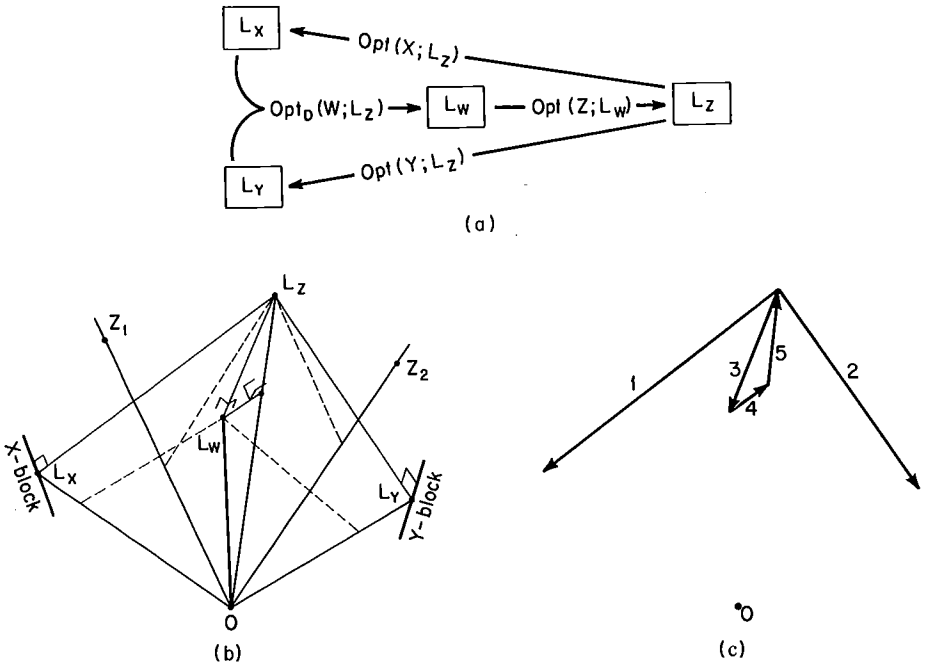
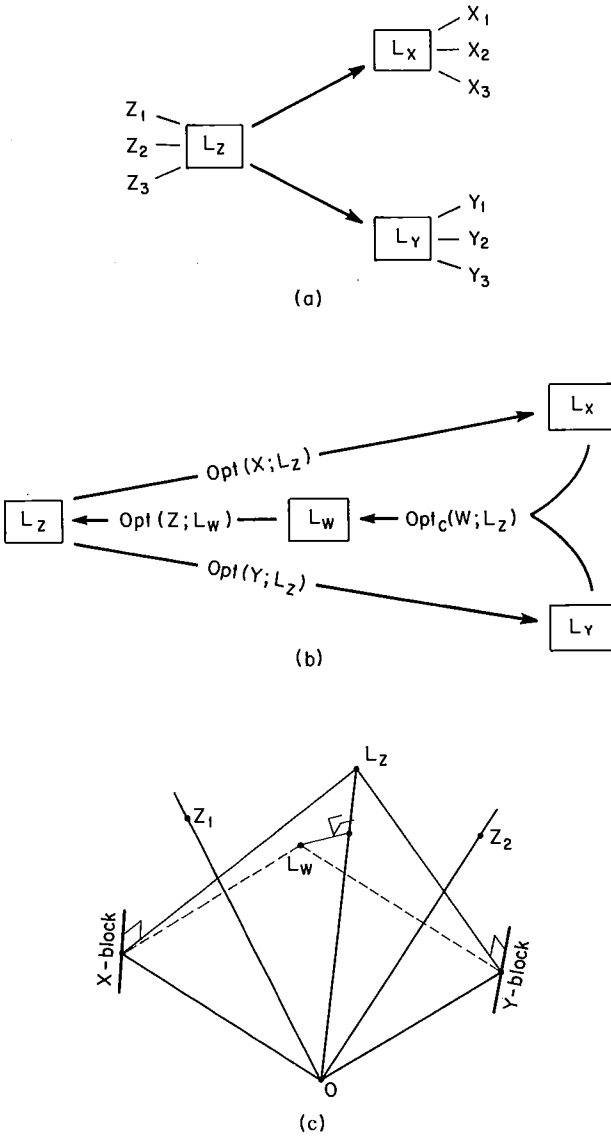


Fig. 9. Recommended algorithm, mode F, for the causal model of Fig. 1. (a) The command diagram displays the ancillary  $L_W$ . (b) The polygon diagram displays the ancillary plane  $W$ . (Compare Fig. 8b.) (c) The sequence of regressions 1-4 along edges of Fig. 9b is followed by a normalization, step 5.



**Fig. 10.** Adjustment of Fig. 9 when the postulated direction of causation is reversed. (a) The revised arrow diagram. (Compare Fig. 1.) (b) The command diagram. Note that  $L_w$  is computed by optimization mode C, rather than mode D as in Fig. 9b. (c) The polygon diagram, incorporating Fig. 8a.

according to mode E instead, with maximand the summed squared *simple* correlations of  $L_Z$  with  $L_X$  and  $L_Y$  separately. There are two dependencies postulated here, corresponding to the two arrows out of the Z-block; in the absence of any instructions to the contrary, their strengths can only be summated.

In a four-stage model, such as that of Adelman and Morriss excerpted in Wold's essay elsewhere in this volume, a suitable command diagram will combine several of these modes as in Fig. 11. Two ancillary structures are invoked here: a plane  $W_1$  spanned by Lvs  $L_E$  and  $L_S$  of the two exogenous blocks "economic levels" and "social conditions"; and a hyper-plane  $W_2$  spanned by those two Lvs and also  $L_P$  ("political conditions") as well—all the Lvs causally prior to the criterion variable "growth rates." Latent variables  $L_S$  and  $L_E$  may be computed (1 and 2 in the figure) from their communicating Lvs  $L_P$  and  $L_G$  by simply averaging of the projections, mode-C optimization, as in Fig. 5. For the operation (4) whose output is  $L_P$ , the instrumental variable in this little model, we additively combine  $\text{Opt}_B(W_1; L_P)$ , the net projection (3) of the exogenous variables onto the P-block (as in Fig. 8a), with the projection of  $L_G$  from the "other direction," causally speaking. (The role of this intermediate block P is the same as that of block Z in Fig. 10a, even though one of the arrows is reversed. The Lv  $L_P$  enters into two different linear determinations the efficacies of which—one a multiple  $R^2$ , one a simple  $r^2$ —should be strictly summated.) Finally, the operation (6) whose output is  $L_G$  invokes the ancillary Lv  $L_{W_2}$  (5) in a

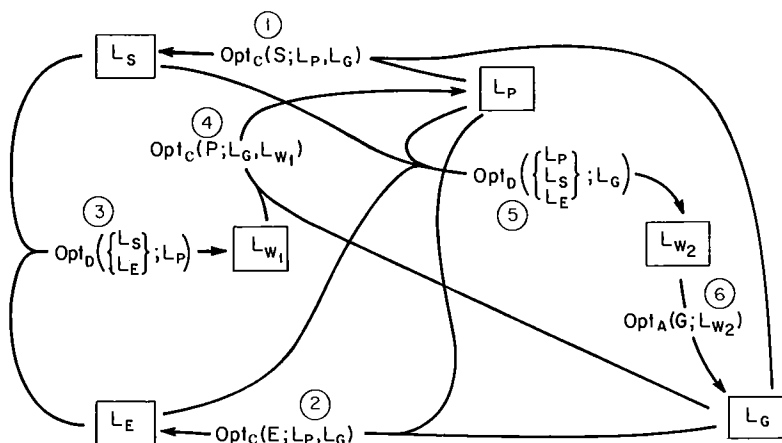


Fig. 11. Command diagram for the Adelman-Morriss model. See Wold's essay in this volume for an explanation of the blocks E, S, P, G. There are two ancillary blocks.  $W_1$  is spanned by Lvs  $L_S$ ,  $L_E$ , and  $W_2$  by Lvs  $L_E$ ,  $L_S$ ,  $L_P$ . The numbering of the Opt commands corresponds to the order of their computation, as described in the text.

construction which extends that of Fig. 8a by one further block. Without information about the matrices  $\Sigma_{EE}$ ,  $\Sigma_{SS}$ ,  $\Sigma_{GG}$ ,  $\Sigma_{PP}$  one cannot guess at their condition numbers or prescribe mode-B projection with respect to any but the blocks  $W_1$ ,  $W_2$  of ancillaries; then operations 1, 2, 4, and 6 are all performed item by item.

## 6. How Soft Modeling Relates to Canonical Analysis and to Maximum-Likelihood Estimation

Expansion of so modest an assumption as the mere block structure into these detailed sequences of regressions can be justified by recourse to the rationale of canonical analysis. Were we presented (well out of context of soft modeling) with the problem of predicting  $Z$ -ness by the  $X$ - and  $Y$ -items in Fig. 1, we would no doubt settle upon a global figure-of-merit, the proportion of variance in  $L_Z$  explained. By this criterion the optimum  $L_Z$  is the first canonical variate of the  $Z$ s with respect to the pool of the  $X$ s and the  $Y$ s together;  $L_X$  would then be the collection of  $X$ -terms in the predictor,  $L_Y$  the  $Y$ -terms. Such an analysis neglects the existence of the  $X$ - or  $Y$ -blocks as separate pools of variables presumed to have some coherence among themselves.

I find the easiest way of declaring the block structure to be by way of side conditions upon the optimization. An insistence that  $L_X$ , the  $X$ -component in the prediction function of  $L_Z$ , be itself an optimal predictor of  $L_Z$  seems to imply the existence of the  $X$ -block satisfactorily. We shall insist, too, that  $L_Y$  be optimal for prediction of  $L_Z$  from the  $Y$ -block alone. That  $L_Z$  which is optimally predicted by a linear combination of  $L_X$  and  $L_Y$  obeying these side conditions is just the fixed-point drawn out in Fig. 9b. In other words, at convergence of the command sequence of Fig. 9a,  $L_Z$  is in canonical relation to the plane of its own best predictors. The soft model, by restating this optimum in terms of partial least squares, has provided a simple iterative algorithm for the analytically unwieldy solution. In all these determinations the polygon diagram lets us see what we are doing.

The preceding paragraph may be thought of as adumbrating a "model" which the partial least squares (PLS) procedure is "estimating." Soft modeling is in effect the estimation of a recursive system of simultaneous linear equations under very specialized nonlinear constraints. Nevertheless, most writers on the subject, in particular Wold and myself, choose not to make much of this interpretation, for two reasons. First, in the context of any family of distributions likely to have generated the data in practice, the constraints make no sense at all. They are not counterfactual, merely irrelevant. In specifying that the analysis be consistent with the only prior



knowledge we have, the block-causal structure, they speak not about the data but about the explanatory use we propose to make of them, explanations whose strength is embodied in the sums-of-squares we are optimizing. Second, from such a model it would be pointless to pass to distribution-based estimates of standard errors in the parameters. These errors conform only to sampling variation in the units of analysis, but PLS is applied rather in the attempt to smooth out into Lvs those long lists of proxies. The sampling variation most crucial for reasoned applications is of those lists of items themselves, and as the items are not drawn randomly from any universe, no sampling theory is of any avail here at all.

Axiomatic simplicity aside, the main advantage of soft modeling by the partial least squares approach over likelihood-based estimators such as LISREL is the proximity of maximands to our intuitive statistical experience, the possibility of viewing the fixed-points in diagrams of perpendiculars and linear combinations. Only during a single regression is likelihood to be seen in the dual space, manifested as ellipses about the foot of a perpendicular. Otherwise, likelihood relates to our vector geometry not by these visible constructions but through hypervolumes and determinants, far too subtly for any simple depiction at all.

In short, likelihood is invisible. Its maxima have no tangible attributes in terms of the subsystems of a model, the arenas wherein our explanatory theories stand or fall. At the conclusion of a partial least-squares analysis, by contrast, we have a collection of true sentences, embodying geometric facts, about precisely those local optima which we know how to interpret: little two- and three-block configurations. In "complex situations with scarce prior information," least-squares optima are treacherous enough, what with outliers, multicollinearity, item selection bias, that ubiquitous misspecification of levels, and all the other familiar difficulties. It seems gratuitous to augment these very real frustrations by the introduction of multivariate normality assumptions, various hypotheses of zero residual correlation, and the like, just for the sake of arriving at asymptotic estimates of the covariance matrix of error of estimate. The decentralization of soft modeling corresponds to a lack of faith, in interdisciplinary problems of any complexity, that the data will bear the imposition of a single function, however general, relating the observed covariance structure to a theoretical distribution.