LEAST SQUARES PATH ANALYSIS WITH OPTIMAL SCALING

Henny Coolen
Section Research Consultancy
Erasmus University
Burgemeester Oudlaan 50
3062 PA Rotterdam
The Netherlands

Jan de Leeuw
Department of Data Theory
Leiden University
Middelstegracht 4
2313 TW Leiden
The Netherlands

Paper to be presented at the Fifth International Symposium Data Analysis and Informatics. Versailles, France. September 29 - October 2, 1987.

INTRODUCTION

The method of path analysis was introduced by Sewall Wright (1921, 1934), who initially applied the method to the field of genetics. In 1925 he published Corn and Hog Correlations which contains a path model of six equations with thirteen variables, some of which were latent. According to Goldberger(1972) this is the first econometric application of structural equation models, at least the first of this size and complexity. It seems that Wright(1931) was also one of the first to publish a social scientific application. Although the title of the article may suggest differently, it contains among others a causal model for the explanation of child's IQ by such factors as heredity, parent's IQ and environment. Although path analysis was extensively used in population genetics, it remained an obscure method in the other sciences for a long time.

Structural equation models were, of course, widely being used by econometricians, who also worked on the solution for such problems as identification, maximum likelihood estimation and correlated errors(Johnston, 1963). But Wright's work was seldom mentioned, although he noticed several of these problems which he solved in an ad hoc way(Goldberger, 1972).

In the sixties other social scientists started to use causal modeling techniques, and this lead to the rediscovery of the work of Wright. Blalock(1964) still leans on the econometric work on structural equation models and mentions Wright only in an appendix. But Duncan(1966) gave an exposition of path analysis that draws heavily upon Wright.

The occupation of sociologists and psychologists with path

analysis lead to some new developments. One of these was the systematic incorporation of latent variables in path models (Hauser and Goldberger, 1971). This meant a combination of the econometric approach, with its focus on the equations, and psychometric theory(especially on factor analysis), with its emphasis on latent variables. It resulted in path analysis methods for the analysis of numerical and latent variables. Joreskog (1973, 1982) introduced a maximum likelihood approach and Wold (1975, 1982) developed a least squares technique for the estimation of these models.

Other developments focused on overcoming the deficient measurement characteristics of many of the variables that are used in the social sciences. In loglinear analysis for instance the discrete character of most of these variables is taken as a starting point(Bishop, Fienberg and Holland, 1975). And the multiway table of the variables in a causal model is analysed and decomposed. This implies that only relatively small models can be analyzed, because of the empty cell problem. Recently Muthen (1983) combined the maximum likelihood models with the discrete approach into one framework.

In this paper an alternative approach and solution is presented. The classical method of path analysis will be combined with the notion of least squares optimal scaling(Gifi, 1981; Young, 1981), which means that non-numerical variables are quantified in a way that is optimal in a least squares sense. By doing so we obtain a very general method of path analysis in which we not only can choose a suitable path model, but within the model we can also choose the measurement level of each of the variables separately.

LEAST SQUARES PATH ANALYSIS

Suppose we deal with n observations on m standardized variables $x_1,\ldots,x_m$. According to some theory these variables form a path model. Path models can be represented in several ways. A popular and attractive way is by means of an arrow diagram. In such a diagram the variables are the corners, drawn as boxes, and the relationships between the variables are the edges, represented as arrows. If there is an arrow from variable x_j to x_k we say that x_j is a direct cause of x_k and variable x_k is called a direct effect of x_j. If there runs a path from x_j to x_1 we shall call x_j a cause of x_1, while x_1 is an effect of x_j. Figure 1 shows a simple path model with four variables.

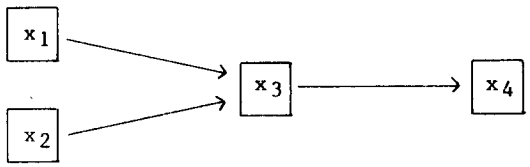

Figure 1. A four-variable recursive path model

Variables at which no arrows arrive are called exogenous, all other variables in the model are endogenous. An important class of path diagrams is transitive, which means that a path starting in a corner never returns to that corner. The diagram in fig. 1 is transitive and the corresponding path model is called recursive. This model would become non-recursive with an arrow from x_4 to x_1 and/or x_2.

A second and more quantitative way of representing path models is by specifying them as a set of linear structural equat-

ions. Such a set of equations is formed by taking each endogenous variable in turn and writing it as a linear function of its direct causes plus an error term. The linear structural model corresponding with fig. 1 is

$$x_3 = b_{31}x_1 + b_{32}x_2 + e_3 \tag{1a}$$

$$x_4 = b_{43}x_3 + e_4 \tag{1b}$$

Suppose E is a subset of the index set $\{1,\ldots,m\}$ that indicates those variables that are endogenous. In least squares path analysis the path coefficients are computed by minimizing the loss function

$$\sigma(B) = \sum_{s\in E}^{p} ||x_s - \sum_{l\neq s}^{m} b_{sl}x_l||^2 \tag{2}$$

$B = \{b_{jk}\}$, the mxp matrix of path coefficients, is restricted by requiring that $b_{jk} = 0$ $(j,k = 1,\ldots,m)$ if x_k is not a direct cause of x_j. Moreover, for recursive path models B will be lower-triangular. Minimization of (2) is quite simple. We solve p linear regression problems, one for each endogenous variable (Wold, 1954). Let b_s be column s of B having k_s unrestricted elements, and let U_s be the nxk_s matrix of the x_j that correspond with the unrestricted part of b_s. Then the solution for b_s is given by

$$b_s = (U_s^T U_s)^{-1} U_s^T x_s \tag{3}$$

So the parameters of a system of equations such as (1) are fitted by single equation ordinary least squares. And the fit of the model is given by the average of the squared multiple correlations.

The exposition of path analysis so far applies to numerical(or quantitative) variables only, because the computation of multiple correlations is somewhat problematical when (some of) the variables are non-numerical. In order to be able to use non-numerical variables in path analysis we will have to code them, but then the size of the multiple correlations will depend on the particular coding chosen. Since in many applications non-numerical variables will occur, we must either use only numerical variables or accept the fact that the solutions will depend on the particular codings chosen. In our approach we shall be looking for codings of non-numerical variables which are optimal in a well-defined sense.

OPTIMAL SCALING

In this paper we shall deal only with variables that can take a finite number of values, very often much smaller than the number of observations. Besides, we shall also only be interested in discrete transformations, by which we mean that observations with the same original score will receive the same transformed value. With these preliminaries in mind we can start our description of optimal scaling.

As stated above in many applications of path analysis, especially but not only in the social sciences, we encounter non-numerical variables. We speak of an ordinal variable when the variable consists of a set of ordered classes. While nominal variables are classifications made up of just a set of equivalence classes. We may call these types of variables partly known.

We know the class or category to which an observation belongs, and for ordinal variables we know the order between the classes too, but we do not know what numerical values to associate with each of the classes.

Sometimes path models even contain variables that are completely unknown(Hauser and Goldberger, 1971). This means that we know the place of the variable in the path model, but we do not have any observations on it. Variables which are unknown in this sense are called latent or unobserved. They can both be endogenous and exogenous. Variables which are not latent are called manifest, regardless of their measurement level. Our approach is to make ordinal, nominal and latent variables suitable for path analysis by optimally scaling these variables.

The general idea behind optimal scaling is to scale the variables in a way that optimizes an objective criterion, which in the case of path analysis would be an appropriate version of loss function (2). A scaling (or quantification, or transformation) of a variable is a real valued function defined on its values. We shall use the notation S_j: x_j -> **R**. The type of scaling that is employed will be determined by what we know or assume about a variable, i.e. which measurement level we associate with a variable.

For latent variables this implies no extra constraints on their scaling, simply because we have no observations for this type of variable. We only require that the scaling of a latent variable conditionally minimizes the loss function.

If we are dealing with a nominal variable we require the transformation of such a variable to maintain the equivalence structure of the original values. Let '<>' be the relation 'has

the same value as', then we can express this restriction as

$$x_{ij} <> x_{kj} \quad => \quad S_j(x_{ij}) = S_j(x_{kj})$$

For ordinal variables we require in addition that the transformations be monotonous with the order of the original values. If '<' denotes the empirical order relation, the additional constraint for ordinal variables becomes

$$x_{ij} < x_{kj} \quad => \quad S_j(x_{ij}) < S_j(x_{kj})$$

For a more elaborate treatment of measurement levels and optimal scaling the reader is referred to Gifi(1981) and Young (1981). The treatment of latent as another, very low, measurement level is due to De Leeuw(1984).

DEFINITION OF PATHALS

As before we start again with a path model containing m standardized variables, which are collected in the nxm matrix X. Beside for every variable in the model we take its assumed measurement level into account too. In least squares path analysis with optimal scaling we now have to minimize the loss function

$$\sigma(B;Y) = \sum_{s\in E}^{p} ||y_s - \sum_{l\neq s}^{m} b_{sl}y_l||^2 \qquad (4)$$

with B as before and Y the nxm matrix with optimally scaled

observations, i.e. $y_j = S_j(x_j)$ $(j = 1,\ldots,m)$. This loss function has to minimized over B and Y under the restrictions

$u^T y_j = 0$

$y_j^T y_j = 1$

$y_j \in C_j \qquad , j = 1,\ldots,m.$

Here u is a nxl vector with only ones. The first condition guarantees that the variables are in deviations from their means, and the second restriction gives every variable a variance equal to one. Although these restrictions seem trivial, they are necessary for obtaining a unique solution. The notation $y_j \in C_j$ indicates that there may be scaling restrictions for variable j.

In this approach to path analysis not only the path coefficients but also the scalings of the variables are parameters that have to be estimated. Because a variable will generally occur in more than one equation we cannot use single equation least squares techniques anymore, the scaling of the variables links the equations. This can be illustrated with the simple structural model (1). If we express the equations in the notation of this paragraph they become

$$y_3 = b_{31}y_1 + b_{32}y_2 + e_3 \qquad (5a)$$

$$y_4 = b_{43}y_3 + e_4 \qquad (5b)$$

We see that variable y_3 occurs in both equations, which are consequently not independent anymore.

Note that a latent variable which occurs in only one equation is not very useful. This equation can always be fitted perfectly and the corresponding term drops out of the loss function.

ALTERNATING LEAST SQUARES

The computer program that implements our method of path analysis minimizes loss function (4) by using an iterative algorithm based on the principle of alternating least squares(ALS) (Young, 1981; Gifi, 1981). This principle involves the partitioning of the parameters into sets: the model parameters B and the scaling parameters Y. We then proceed to minimize the loss function by alternatingly optimizing it with respect to one of the subsets. At each stage of the algorithm this gives the conditional least squares estimates of one of the subsets, while keeping all the other parameters fixed at their current value. Once we have obtained the conditional least squares estimates of a subset we replace the old estimates of these parameters by the new ones. We then switch to another set of parameters and repeat the process. By cycling through the sets of parameters in this way we obtain a convergent algorithm. In the PATHALS problem each variable defines a set of scaling parameters and each path equation defines a subset of model parameters. This means that in minimizing (4) two basic subproblems have to be solved.

SUBPROBLEMS

Estimating the path coefficients for fixed scalings of the variables is again done by single equation ordinary least squares. This means that for every equation of the path model we have to solve

$$\left|\left| y_s - \sum_{l \neq s}^{m} b_{sl} y_l \right|\right|^2 \quad , \quad s \in E \qquad (6)$$

for the unrestricted part of b_s, with b_s as before. Let U_s now be the nxk_s matrix of the y_j that correspond with the unrestricted elements of b_s. The solution is

$$b_s = (U_s^T U_s)^{-1} U_s^T y_s \tag{7}$$

This result clearly gives the conditional least squares estimates of the path coefficients.

Finding a solution for the scaling parameters is a bit more complicated. Observe that loss function (4) can be rewritten as

$$\sigma(B;Y) = \sum_{j=1}^{m} \| y_j - \sum_{k \neq j} c_{jk} y_k \|^2 \tag{8}$$

In (8) every variable in the path model, whether endogenous or exogenous, is expressed as a linear combination of those variables with which it is connected through an arrow in the path diagram. The second subproblem can now be formulated as follows. For every variable y_j minimize for fixed scalings of the other variables $y_1,\ldots,y_{j-1},y_{j+1},\ldots,y_m$ and fixed path coefficients B

$$\| y_j - \sum_{k \neq j} c_{jk} y_k \|^2 \quad , \quad j = 1,\ldots,m \tag{9}$$

subject to the constraints

$u^T y_j = 0$

$y_j^T y_j = 1$

$y_j \in C_j$.

Problem (9) is a normalized cone regression problem(Gifi, 1981), that can be solved by first computing an unnormalized solution and subsequently normalizing it(De Leeuw, 1977). The

unnormalized cone regression problem involves two steps. First we minimize the loss with respect to y_j unrestricted. This gives $\tilde{y}_j$. Then we perform the appropriate type of regression of $\tilde{y}_j$ on x_j. The type of regression depends on the measurement level that is imposed on a variable. With ordinal restrictions we do monotone regression(Kruskal, 1964; Gifi, 1981). With nominal variables identity regression is performed(Gifi, 1981). Latent variables do not enter this substep. All variables are subsequently standardized.

THE ALGORITHM

The computer program that implements the iterative algorithm is called PATHALS. It is written in SAS/IML. The programs computes an initial and final solution using the same basic algorithm, which consists of the following main steps.

Iterate until $\sigma(B;Y)$ fails to decrease;

For j = 1 to m Do;

$\tilde{y}_j = \Sigma_{l \neq j} c_{jl} y_l$;

$\hat{y}_j = OS(\tilde{y}_j)$;

$y_j = \hat{y}_j (\hat{y}_j^T \hat{y}_j)^{-1/2}$;

End For Loop;

For s = 1 to p Do;

$b_s = (U_s^T U_s)^{-1} U_s^T y_s$;

End For Loop;

End Iteration Loop;

The first For Loop performs the optimal scaling(OS) of the variables. The unrestricted update of y_j is denoted as $\tilde{y}_j$. The optimally scaled $\tilde{y}_j$ is indicated as $\hat{y}_j$, while y_j is the ultimate standardized update. In the other For Loop the path coefficients are computed. The iteration process is stopped as soon as $\sigma(B;Y)$ decreases by less than a predetermined criterion.

For the computation of the initial solution all manifest variables are treated as numerical and the latent variables are initialized with random values. The program then proceeds to compute the parameters of the latent variables and the path coefficients iteratively. If the path model does not contain latent variables only the second For Loop is executed once in this phase. It immediately gives the initial solution, which in this case is a solution to loss function (2) given the original codings of the variables.

The initial solution serves as the starting point for the computation of the final solution. For this solution the imposed scaling restrictions on the variables are taken into account. The result of this phase consists of the path coefficients and the optimal scalings of the variables. The global convergence of the algorithm has been shown by De Leeuw(1986).

MISSING DATA

For the computation of the initial solution the missing observations on each variable are given the average value of the non-missing observations.

There are many ways of treating missing data according to

optimal scaling theory(Gifi, 1981), two of which have been implemented in the program. If on chooses missing multiple the missing observations on a variable are treated as unique and every missing score gets its own quantification. On the other hand when the missing observations are treated as equivalent they will receive the same quantification. We call this option missing single.

For the computation of the parameters the inclusion of missing data has minor consequences. The missing data do not enter into the optimal scaling step of the first For Loop, but are treated the same way as the non-missing observations in the other two substeps.

COMPARISON WITH SOME OTHER APPROACHES

The partial least squares(PLS) approach to path analysis of Wold(1975, 1982) has some things in common with PATHALS, but also differs from it in some essential respects. What the two techniques have in common is that they are both based on least squares. Moreover, the algorithms of the ALS and PLS programs are very similar. A good description of the principles underlying the PLS or NIPALS algorithms can be found in Wold(1966). It basically comes down to the computation of the parameters by iteratively cycling through non-overlapping sets of parameters.

But these similarities introduce at the same time one of the main differences between the two approaches. In the ALS approach one single global criterion, loss function (4), is optimized. On the other hand several separate least squares solutions are

computed in PLS. Every path model is divided into a measurement model for the computation of the latent variables, and a structural model in which only the path coefficients are the parameters. Both models are fitted separately, which results in least squares solutions for the subproblems. But this does not guarantee, of course, that these solutions are optimal in an overall sense. On the contrary, this will generally not be the case(Dijkstra, 1981; Wold, 1981).

A second important difference occurs by the fact that in PATHALS the measurement levels of the variables are systematically incorporated by applying optimal scaling theory. In the LVPLS version(Lohmoller and Wold, 1982) the treatment of categorical variables seems less systematic. Every category of such a variable becomes a dichotomous variable in the measurement model. And the weights that are obtained by estimating this model are used as the quantifications of the variables.

Another popular method for the estimation of path models is LISREL(Joreskog, 1982; Joreskog and Sorbom, 1984). Its emphasis is on numerical and latent variables, and on estimation by means of maximum likelihood. In its latest version the program contains options for least squares solutions. This makes a certain treatment of ordinal variables possible. The suggestion is to compute polychoric or polyserial correlations instead of product moment correlations when ordinal variables are involved, and to analyze the in this way obtained correlation matrix by means of unweighted least squares(Joreskog and Sorbom, 1984).

Algorithmically the ALS and ML approach can hardly be compared. The LISREL program computes its solution on the basis of the Fletcher-Powell version of the Raphson-Newton method. Al-

though in LISREL the distinction between measurement and structural model is also made, it does estimate the parameters of both simultaneously by optimizing the global likelihood function.

AN EXAMPLE

Our illustration of the PATHALS technique refers to a so called MIMIC model(Joreskog and Goldberger, 1975). In MIMIC models there are two sets of manifest variables. The exogenous variables influence the endogenous variables through the mediation of on or more latent variables. Figure 2 shows an arrow diagram for a simple MIMIC model. Latent variables are drawn as circles in stead of squares.

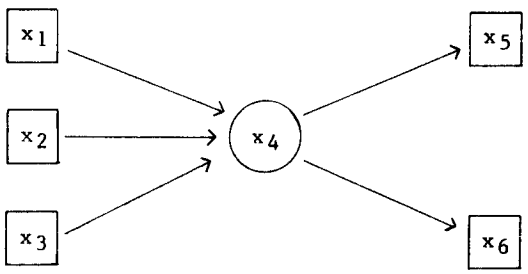

Figure 2. A simple MIMIC model

We used ecological data taken from Van der Aart and Smeenk-Enserink(1975), who reported abundance data for 12 species of hunting spiders in a dune area in the Netherlands. A total of 28 sites were studied, which were also described in terms of a number of environmental variables. We have used a selection of 6 of these made by Ter Braak(1985):

WC Water content, percentage dry weight,

BS Percentage bare sand,

CM Percentage covered by moss layer,

LR Reflection of soil surface at cloudless sky,

FT Percentage covered by fallen leaves or twigs,

CH Percentage covered by herbs layer.

Ter Braak made these variables discrete into 10 categories, coded as 0 through 9.

The results of a MIMIC analysis with two latent variables are given in table 1. The environmental variables are exogenous

	Weights			Residual variances	
	linear		monotone	linear	monotone
WC	-.82	.10	-.96 -.20		
BS	-.06	.11	-.56 .39		
CM	.13	.21	-.14 -.33		
LR	-.02	.56	.28 .12		
FT	.72	-.24	.22 -.15		
CH	-.29	.10	-.71 .43		
S1	-.79	-.09	-.89 .22	.39	.21
S2	.04	-.79	.30 -.87	.36	.21
S3	-.85	-.35	-.88 -.16	.22	.16
S4	-.95	-.10	-.99 .21	.13	.04
S5	-.97	-.06	-.99 .22	.08	.04
S6	-.91	-.13	-.95 .19	.21	.10
S7	-.93	-.48	-.98 .01	.07	.04
S8	-.77	-.11	-.85 .00	.43	.27
S9	-.36	.52	.74 -.48	.53	.32
S10	.18	.88	.07 .90	.25	.16
S11	.52	.71	.48 .71	.36	.18
S12	.53	.53	.57 .54	.54	.31

Table 1. Hunting spider data: linear and monotone solutions

and the species, denotes as S1,...,S12, endogenous. The results of a linear analysis, using the codings of Ter Braak, and of an analysis that computed optimal monotone transformations are both depicted.

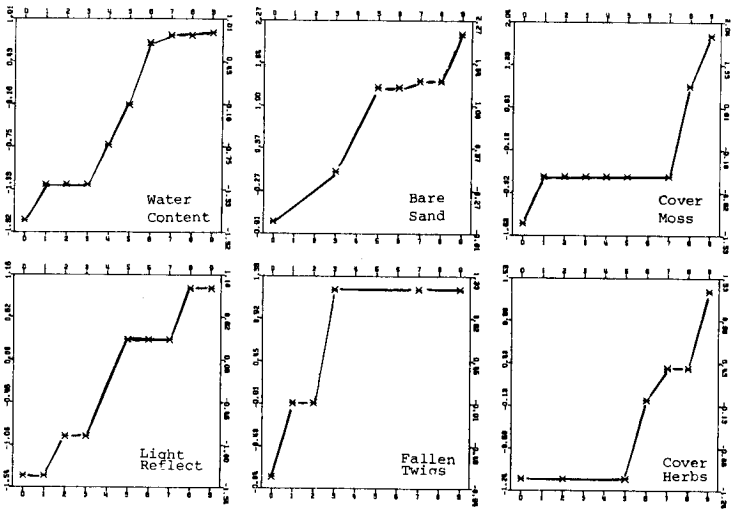

Figure 3. Hunting spider data: optimal monotone transformations

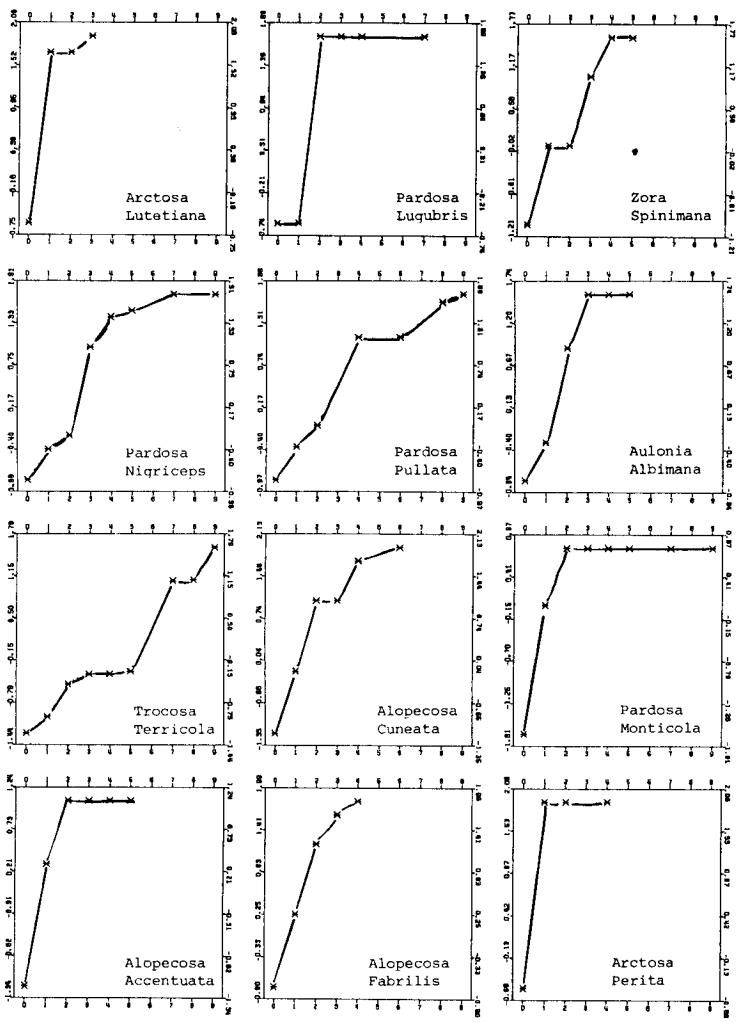

Figure 3 shows the optimal transformations of the variables. The X-axis contains the original category numbers and the Y-axis the optimal scalings. We see a large variety of shapes, roughly linear, two-step, convex, and so on. We shall not give a detailed analysis here, but merely point out some technical aspects and compare the two solutions.

We see from table 1 that the residual variances of the abundance variables decrease considerably. This is also the case for the latent variables, which are not given in the table. For the linear analysis the residuals are .06 and .14, while for the monotone analysis they are .01 and .01. This means that they are almost completely 'explained' by the monotone transformed environmental variables. The interpretation of the latent variables is facilitated by correlating them with the transformed variables. If we do this we find that the first latent variable correlates -.75 with both WC and CH, while the second one correlates .80 with LR and -.80 with FT.

The analysis shows the advantage of the PATHALS technique. By optimally scaling the variables we need fewer dimensions to account for a large proportion of the variance, because much of the remaining variation after a linear analysis is taken care of by the transformations. Beside, the optimal transformations often give useful additional information about the data.

DISCUSSION

The method of path analysis described in this paper is an extension of classical least squares path analysis. The data

analyst can now not only choose an appropriate path model, but also within the model for each variable a suitable class of transformations from which an optimal one must be chosen. This enhances the applicability of path analysis. At the same time this can lead to new problems. The scaling of the variables creates new parameters which have to be fitted, while there are fewer restrictions on the data. This may lead to chance capitalization and triviality, dangers against which we have to guard ourselves. It is necessary to investigate the stability of the solutions, and we must also be able to give interpretations for the results.

The PATHALS technique and program, as described here, will soon be provided with some new options. One of these shall comprise the possibility of a more continuous treatment of numerical data. In the hunting spider example, for instance, the continuous environmental variables were made discrete before they could be analyzed. This means that we ignored the prior information that these variables were continuous. Consequently all observations within an interval got the same quantification and the fact that the intervals are connected was ignored. Our theory can be generalized in such a way that it will take the smoothness of transformations into account. A fundamental role in this extension is played by the 'B-spline basis'(cf. De Leeuw, Van Rijckevorsel, Van der Wouden, 1981; Van Rijckevorsel, 1982). Computationally this new option does not create any trouble. We must introduce a new subproblem into the alternating least squares cycles of PATHALS(cf. Coolen, Van Rijckevorsel, De Leeuw, 1982).

REFERENCES

Bishop, Y.M.M., S.E. Fienberg and P.W. Holland (1975) Discrete multivariate analysis: theory and practice. Cambridge, Mass., M.I.T. Press.

Blalock, H.M. (1964) Causal inference in non experimental research. Chapel Hill, University of North Carolina Press.

Coolen, H.C.C.H., J. van Rijckevorsel and J. de Leeuw (1982) An algorithm for nonlinear principal components analysis with B-splines by means of alternating least squares. in: Caussinus H., P. Ettinger and J.R. Mathieu (Eds.) (1982) Compstat 1982, part II. Wien, Physika-Verlag.

De Leeuw, J. (1977) Normalized cone regression. Research Report, University of Leiden, Department of Datatheory, The Netherlands.

De Leeuw, J. (1984) Least squares and maximum likelihood for causal models with discrete variables. Report RR-84-09, University of Leiden, Department of Datatheory, The Netherlands.

De Leeuw, J. (1986) Multivariate analysis with optimal scaling. Report RR-86-01, University of Leiden, Department of Datatheory, The Netherlands.

De Leeuw, J., J. van Rijckevorsel and H. van der Wouden (1981) Nonlinear principal components analysis with B-splines. Methods of Operations Research, 33, 379-393.

Dijkstra, T.K. (1981) Latent variables in linear stochastic models. Doctoral Dissertation, University of Groningen, The Netherlands.

Duncan, O.D. (1966) Path analysis: sociological examples. American Journal of Sociology, 72, 1-16.

Gifi, A. (1981) Nonlinear multivariate analysis. University of Leiden, Department of Datatheory, The Netherlands.

Goldberger, A.S. (1972) Structural equation methods in the social sciences. Econometrica, 40, 979-1001.

Goldberger, A.S. and O.D. Duncan (Eds.) (1973) Structural equation models in the social sciences. New York, Seminar Press.

Hauser, R.M. and A.S. Goldberger (1971) The treatment of unobservable variables in path analysis. in: Costner, H.L. (Ed.) Sociological methodology 1971. San Francisco, Jossey-Bass.

Johnston, J.J. (1963) Econometric methods. New York, McGraw Hill.

Joreskog, K.G. (1973) A general method for estimating a linear structural equation system. in: Goldberger and Duncan(1973)

Joreskog, K.G. (1982) The LISREL approach to causal model-building in the social sciences. in: Joreskog and Wold (1982).

Joreskog, K.G. and A.S. Goldberger (1975) Estimation of a model with multiple indicators and multiple causes of a single latent variable. Journal of the Americam Statistical Association, 70, 631-639.

Joreskog, K.G. and D. Sorbom (1984) LISREL VI: analysis of linear structural relationships by the method of maximum likelihood. Chicago, International Educational Services.

Joreskog, K.G. and H. Wold (1982) Systems under indirect observation: causality, structure, prediction. Amsterdam, North Holland Publishing Company.

Kruskal, J. (1964) Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika, 29, 1-27.

Lohmoller, J.B. and H. Wold (1982) Introduction to PLS estimation of path models with latent variables, including some recent developments on mixed scales variables. Munchen, Hochschule der Bundeswehr.

Muthen, B. (1983) Latent variable structural equation modeling with categorical data. Journal of Econometrics, 22, 43-67.

Ter Braak, C.L.F. (1985) Canonical Correspondence Analysis. Wageningen: Institute TNO for Mathematics, Information processing and Statistics, The Netherlands.

Van der Aart, P.J.M. and N. Smeek-Enserink (1975) Correlation between distributions of hunting spiders(Lycosidae, Ctenidae) and environmental characteristics in a dune area. Netherlands Journal of Zoology, 25, 1-45.

Van Rijckevorsel, J. (1982) Canonical analysis with B-splines. in: Caussinus H., P. Ettinger and R. Tomassone (Eds.) (1982) Compstat 1982, part I. Wien, Physika-Verlag.

Wold, H. (1954) Causality and econometrics. Econometrica, 22, 162-177.

Wold, H. (1966) Nonlinear estimation by iterative least squares procedures. in: David, F.N. (Ed.) (1966) Research papers in statistics: festschrift for J. Neyman. New York, John Wiley.

Wold, H. (1975) Path models with latent variables: the NIPALS approach. in: Blalock, H.M., F.M. Borodkin, R. Boudon and V. Capecchi (Eds.) (1975) Quantitative sociology. New York, Seminar Press.

Wold, H. (1981) Model construction and evaluation when theoretical knowledge is scarce: on the theory and applications of partial least squares. in: Kmenta, J. and J. Ramsey (Eds.) (1981) Model evaluation in econometrics. New York, Academic Press.

Wold, H. (1982) Soft modeling: The basic design and some extensions. in: Joreskog and Wold (1982).

Wright, S. (1921) Correlation and causation. Journal Agricultural Research, 20, 557-585.

Wright, S. (1925) Corn and hog correlations. Washington: United States Department of Agriculture, Bulletin 1300.

Wright, S. (1931) Statistical methods in biology. Journal of the Americam Statistical Association, Supplement, 26, 155-163.

Wright, S. (1934) The method of path coefficients. Annals of mathematical statistics, 5, 161-215.

Young, F.W. (1981) Quantitative analysis of qualitative data. Psychometrika, 46, 347-388.