

## Factor analysis of causal models

Th. J. Ferrari and J. Mol

Institute for Soil Fertility, van Hallstraat 3, Groningen, The Netherlands;  
Institute for Economic Research, Oude Boteringstraat 44, Groningen, The Netherlands

Received 16 May, 1966

### Summary

A number of observations was constructed artificially by giving the independent variables of a system of equations various values; in this way the values of the dependent variables could be computed. Next these data have been analysed by factor analyses of which the results were compared with the content of the original model.

It appears, that the number of aspects is equal to the number of independent variables. It is very difficult to interpret the results of these analyses without rotation. This rotation must be performed according to the *caumax method*: the possibility to a causal interpretation is maximized by a rotation in which every reference axis is rotated as much as possible upon an independent variable; the dependent variables are not taken into account. By this an identification of the factor analytical model is obtained. The results are not altered very greatly by error factors and correlations between the independent variables.

### Introduction

We know a number of methods to describe and to analyse the complex relationships in certain fields of investigation such as economics, psychology, sociology and agriculture; for this purpose factor analysis and path coefficient analysis or simultaneous equations analysis are used widely. Purpose and models of both those analyses differ rather greatly (Ferrari, 1965). The latter is used mainly in order to test more or less developed hypotheses and to quantify the parameters of them. The factor analysis can also be used for this purpose, in a lesser degree, however; the analysis can be applied better for drawing up or for correcting hypotheses (Cattell, 1965).

The model on which factor analysis is based, is little specified. Neither is the system identified, by which the solution is not unique and rotations of the reference axes are allowed and necessary (Ferrari, 1965). In connection with this *underidentification* the interpretation of the analytical results meets mostly with many difficulties. The researcher understands with difficulty the meaning of the results obtained (Liberg and Mol, 1961).

It is in the nature of things, that the results of the method with the simultaneous equations are easier to understand. In connection with this a comparison between the results of both methods of analysis applied on the same data may be meaningful. Mol (1966) already demonstrated the significance of the number of exogeneous variables in an economic model for the results of a factor analysis. In the following we shall investigate by means of factor analysis the conduct and relationships of variables

in which investigation the observational data are computed after giving one or more primary factors (exogeneous or independent variables) of a path coefficients model a range of various values. We shall compare the results of such analyses with the model from which the data originated. These comparisons will give the possibility to draw some rules for making more specified models and for the study of the results (by rotations).

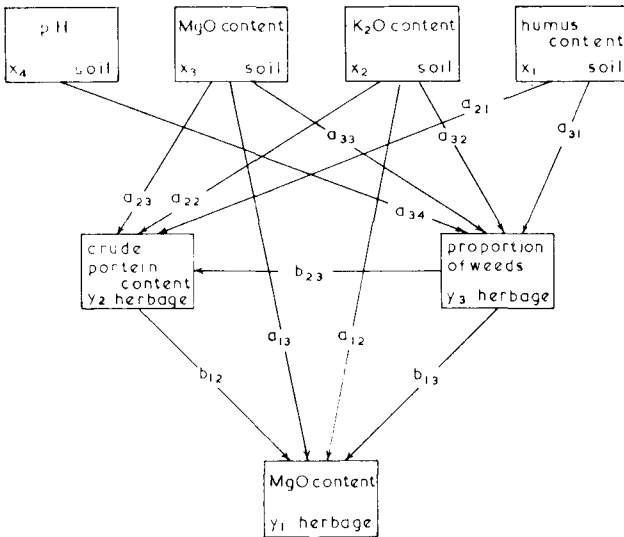


Fig. 1 Model of the influences of 4 independent variables (primary causes) on 3 dependent variables (effects).

The data are obtained from an investigation into the influences of some variables on the magnesium content of herbage, described before (Ferrari, 1964). This model is given in Fig. 1. The model shows how the variables: humus content (x<sub>1</sub>), magnesium content (x<sub>3</sub>), potash content (x<sub>2</sub>) and pH of the soil (x<sub>4</sub>) influence directly and indirectly via changes of the proportion of weeds (y<sub>3</sub>) and of the crude protein content (y<sub>2</sub>) the magnesium content of the herbage (y<sub>1</sub>). This model can be described by the following structural equations<sup>1</sup>:

$$y_3 = a_{31}x_1 + a_{32}x_2 + a_{33}x_3 + a_{34}x_4 \quad (1)$$

$$y_2 = a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + b_{23}y_3 \quad (2)$$

$$y_1 = a_{12}x_2 + a_{13}x_3 + b_{12}y_2 + b_{13}y_3 \quad (3)$$

Terms x<sub>i</sub> in the equations are the primary causes or the exogeneous variables of the economics. The changes of the effects y<sub>i</sub> caused by the changes of these independent variables can be computed by the equations in the reduced form; in each reduced

<sup>1</sup> In these and following equations the level (constant term) will be left out of consideration.

structural equation an effect is expressed as a function of independent variables only. We used for the calculations the parameter values of Table 1, found in an earlier research (Ferrari, 1963).

Next we investigate on which way the effects in this model change, when only cause  $x_3$  is changed and the other causes remain constant. We shall also examine the results of changes of two independent variables ( $x_3$  and  $x_2$ ) at constant values of  $x_1$  and  $x_4$ . Finally we shall look at the influences of changes of three or more independent variables. In this way we simulate the response of the nature to certain changes. We shall apply factor analyses to the data computed, the results of which will be compared with the original model. From these comparisons we can make some inferences.

Table 1 Parameter values of the equations 1, 2 and 3

$a_{31} = 1.67$	$a_{21} = -0.74$	$a_{12} = -0.0038$
$a_{32} = -0.23$	$a_{22} = 0.11$	$a_{13} = 0.0004$
$a_{33} = -0.031$	$a_{23} = 0.011$	$b_{13} = 0.0041$
$a_{34} = 5.26$	$b_{23} = 0.20$	$b_{12} = 0.0083$

### Response of the dependent variables to changes of one or more independent ones

We give the variable  $x_3$  a number (20) of various values chosen at random but not lying outside the range of original values. Next the values of  $y_3$ ,  $y_2$  and  $y_1$  belonging to the twenty values of  $x_3$  must be calculated. As we have to do in this case with changes of  $x_3$  only, we need a simplified system of equations, viz.:

$$y_3 = a_{33}x_3 \tag{4}$$

$$y_2 = a_{23}x_3 + b_{23}y_3 \tag{5}$$

$$y_1 = a_{13}x_3 + b_{12}y_2 + b_{13}y_3 \tag{6}$$

In this system equation 4 has already the reduced form. The reduced form of the other equations is as follows:

$$y_2 = (a_{23} + a_{33}b_{23}) x_3 \tag{5a}$$

$$y_1 = (a_{13} + a_{23}b_{12} + a_{33}b_{12} + a_{33}b_{13}) x_3 \tag{6a}$$

Under the assumption that nature will respond according to the parameter values of Table 1, the equations become:

$$y_3 = -0.031 x_3 \tag{4b}$$

$$y_2 = 0.0048 x_3 \tag{5b}$$

$$y_1 = 0.0003127 x_3 \tag{6b}$$

Using these equations we can now compute the values of  $y_3$ ,  $y_2$  and  $y_1$  for every of the twenty values of  $x_3$ . Table 2 gives some results.

In general a researcher will be ignorant of the relationships governing these numbers. It is his task to discover these relationships. However, we know these relationships

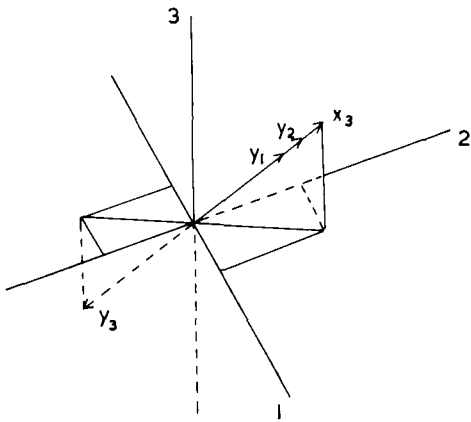


Fig.2 Vector configuration for 1 independent variable.

and a simple computation shows us, that the numbers of the second row equal  $-0.031$  times the corresponding numbers of the first row; the numbers of the  $y_2$ -row equal  $-0.155$  times the numbers of the second row etc. Consequently the rows (and the columns) of this table are dependent on each other; if the numbers of one row are given, the numbers of the other rows can be computed by these ratios. This means in mathematical terms, that the rank of this matrix equals 1.

Table 2 Some computed observational data, with 1 independent variable

Variable	Observation				
	1	2	3	.....	20
$x_3$	167	122	50	.....	185
$y_3$	-5.18	-3.78	-1.55	.....	-5.74
$y_2$	0.80	0.59	0.24	.....	0.88
$y_1$	0.052	0.038	0.016	.....	0.058

For a good understanding of the results of a factor analysis the geometric representation of the idea of rank may be significant. Starting point is the representation of the correlations between the four variables into a 20-dimensional space with rectangular coordinates, in which frame the twenty observations are represented as reference axes. The variables themselves are represented as *vectors* in this system. In Fig. 2 the axes of only three observations of Table 2 are drawn; in reality the four vectors are contained in a 20-dimensional space. It appears in this case, that the four vectors have the same or the opposite direction; these directions coincide. This means, that the correlations are complete and the correlation coefficients are  $+1$  or  $-1$ . The positions of these four vectors can be described with a one-dimensional space. This one-dimensional vector configuration corresponds to the one-dimensional space in the well-known correlation representation in which the variables are taken as reference axes.

The fact the variation of the three dependent variables  $y_3$ ,  $y_2$  and  $y_1$  has been caused by only one independent variable, makes that the vector configuration has a dimension of one; the space of the joint movements of the variables has the same dimension of one.

Further we notice, that the actual changes of the effect variables are described by three equations. The space of the vector configuration informs of the number of equations, but not of the nature of these ones. The researcher, having determined this number by a factor analysis, has the liberty to give divergent explanations to this movement pattern with different equation systems.

Next we change independently of each other the variables  $x_3$  and  $x_2$  of the model. The variables  $x_1$  and  $x_4$  remain constant and can be left out of consideration in equation 1, 2 and 3. The reduced equations are now as follows:

$$y_3 = a_{33}x_2 + a_{33}x_3 \tag{7a}$$

$$y_2 = (a_{22} + a_{32}b_{23}) x_2 + (a_{23} + a_{33}b_{23}) x_3 \tag{8a}$$

$$y_1 = (a_{12} + a_{22}b_{12} + a_{32}b_{23}b_{12} + a_{33}b_{13}) x_2 + (a_{12} + a_{22}b_{12} + a_{33}b_{23}b_{12} + a_{33}b_{13}) x_3 \tag{9a}$$

We assume again, that the model responds to these changes according to the data of Table 2 and obtain the following equations:

$$y_3 = -0.230 x_2 - 0.031 x_3 \tag{7b}$$

$$y_2 = 0.0640 x_2 + 0.0048 x_3 \tag{8b}$$

$$y_1 = -0.00421 x_2 + 0.00031 x_3 \tag{9b}$$

We use the equations in order to compute the values of  $y_3$ ,  $y_2$  and  $y_1$  for every of the twenty values of  $x_3$  and  $x_2$ . Table 3 gives some results.

Table 3 Some computed observational data, with 2 independent variables

Variable	Observation				
	1	2	3	.....	20
$x_3$	167	122	50	.....	185
$x_2$	21	15	21	.....	6
$y_3$	-10.01	-7.23	-6.38	.....	-7.12
$y_2$	2.15	1.55	1.58	.....	1.27
$y_1$	-0.037	-0.025	-0.073	.....	0.032

These data have been produced by changes of two independent variables only; from this it follows, that the rank of the matrix of data is 2. The numbers of each row can be computed, if the numbers of any two rows are given. The correlations between the numbers in the rows (variables) can also be studied by the method just described. These relationships are drawn in Fig. 3 in which the three reference axes (observations) are left out for reasons of drawing technique. Now the vectors (variables) appear not to be contained in a one-dimensional space, but to be spread out in a two-dimensional plane. The variables  $x_3$  and  $x_2$  are not correlated and therefore the

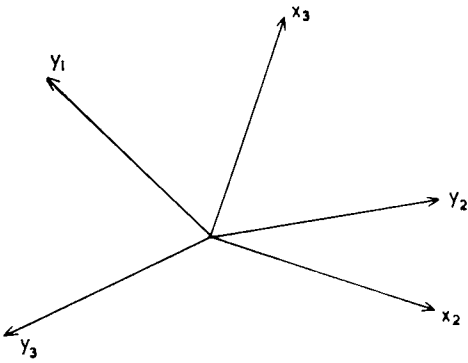


Fig. 3 Vector configuration for 2 independent variables.

directions of these vectors are mutually orthogonal. The variable  $y_2$  is correlated positively with the variables  $x_2$  and  $x_3$ , the variable  $y_3$  on the contrary negatively. Finally  $y_1$  is correlated negatively with  $x_2$  and positively with  $x_3$ .

As the relationships between the variables (*joint movement pattern*) of the first example with one independent variable could be represented by a one-dimensional space, we now need a two-dimensional space to describe the relationships produced by two independent variables. The rank two corresponds to a two-dimensional movement pattern. The researcher ignorant of the creative model can now conclude, that the number of independent variables has been two and that the number of equations describing the model equals to the total number of variables minus the number of independent variables or dimensions of the vector configuration. But he has again the freedom to interpret these relationships and equations in various ways.

In analogy to the results discussed above, it is clear that we shall get a three-dimensional vector configuration by the changes of three independent variables etc. Therefore we can formulate the following rule:

*The number of equations in a model is equal to the number of variables minus the dimensions of the vector configuration; this dimensional number equals the number of independent variables.*

With the aid of this rule the researcher is able to investigate whether the assumed model he wants to test is in accordance with reality. Besides, he can use his knowledge of the dimensions of the movement pattern and of the variables, joined with each other in a certain movement direction, in order to formulate a more specified hypothesis. The rule can also be used to give a better interpretation of the results of factor analytic research.

### Factor analysis and rotation

The factor analysis tries to determine the dimensions of the vector configuration and to indicate in which measure the variables in a certain aspect (dimension) are related with each other. The analysis represents the vector configuration as a projection upon a rectangular axes frame (reference axes) with the same origin as the vectors. The positions of this reference frame with respect to the vector configura-

tion depends upon the solution method<sup>2</sup> and upon the variables taken up (Harman, 1960; Seal, 1964). The factor analytic model is not identified (Ferrari, 1964), so that the solution obtained is not unique. By rotation of the reference axes each solution can be reached in which the vector configuration and the number of dimensions of the movement pattern remain unchanged. This underidentification and the many solutions connected herewith make the interpretation of the analytic solution difficult. A comparison between the original model and the results of the factor analyses applied to data produced by this model points to the possibility to reach an identification by a certain rotation of the reference axes. This comparison is also useful to indicate the potentialities and limits of the factor analysis.

We start with an analysis of the data originated from changes of two independent variables  $x_2$  and  $x_3$  (Table 3); the changes themselves of  $x_2$  and  $x_3$  are not correlated with each other. The researcher is supposed to be ignorant of this background and tries to get some information about the origin of the relationships between the variables by a factor analysis. For this purpose he intercorrelates the variables (Table 4).

Table 4 Correlation coefficients, computed with data of the complete Table 3

	$x_2$	$x_3$	$y_3$	$y_2$	$y_1$
$x_2$	1.00	0	-0.74	+0.89	-0.89
$x_3$		1.00	-0.68	0.47	0.44
$y_3$			1.00	-0.96	0.35
$y_2$				1.00	-0.59
$y_1$					1.00

Table 5 Principle-factor solution; model with 2 independent variables

Variable	Aspect		
	$F_1$	$F_2$	Sum $a_{im}^2$
$x_2$	0.94	-0.35	1.01
$x_3$	0.35	0.94	1.01
$y_3$	-0.92	-0.38	1.00
$y_2$	0.99	0.12	0.99
$y_1$	-0.68	0.73	1.00
Eigenvalue	3.30	1.70	5.00

This matrix of correlation coefficients needs little comment as the background is known to us; the absence of correlation between the independent variables  $x_2$  and  $x_3$  is self-evident, we have varied these variables independently of each other. We are interested more in the results of a factor analysis applied to this matrix (Table 5). The coefficients  $a_{im}$  are a measure of the degree of correlation between aspect and variables; they vary from +1 to -1. The square of a coefficient times 100 indi-

<sup>2</sup> We have always used the principal factor method of Hotelling.

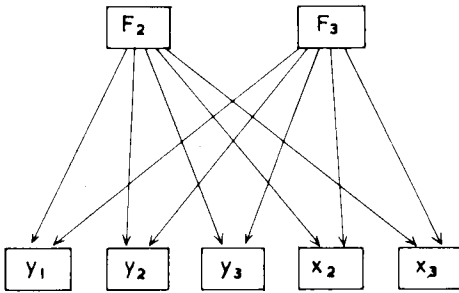


Fig. 4 Factor analytic model with 3 dependent and 2 independent variables.

ates the proportion of the total variance of the variable connected with the aspect concerned; the sum of these squares therefore equals 1 (total variance). The original correlation coefficient between two variables can be reproduced from this aspect matrix by the sum of the products of pairs of corresponding numbers in the two rows concerned (inner product), so the correlation coefficient  $r_{23}$  between the variables  $x_2$  and  $x_3$  is equal to  $0.94 \times 0.35 + (-0.35 \times 0.94) = 0$  etc. Finally the eigenvalue is a measure for the total variance accounted for by the corresponding aspect. A comparison between the model of Fig. 1 without the variables  $x_1$  and  $x_4$  and the model of the aspect analysis (Fig. 4) shows, that we may not expect to obtain back the original information out of the factor analysis. The original model is characterized by chain processes and these are missing in the factor analytic model. The model with the path coefficients contains information about the sizes of the separate influences, and the possibility to achieve these parameters is absent in the factor analysis. The analysis gives information about the simultaneous changes of the variables, connected with the aspects. These changes are expressed as proportions of the total variance: 88 % of the  $x_2$ -variance is connected with 12 % of the  $x_3$ -variance, with 85 % of the  $y_3$ -variance etc. The same applies to the data of aspect 2; in this aspect the variances of the variables independent of those of aspect 1 are represented.

The data of Table 5 show that the variances of all variables is accounted for entirely by both aspects; the sums of squares always equal 1. This is according to the rule: two independent variables must give two aspects or movement patterns.

The correlations between an aspect and the independent variables  $x_2$  and  $x_3$  are contrary to our expectations. For we varied these variables independently from each other, as is expressed by the zero-correlation coefficient. However, we have a situation in which both variables vary simultaneously in the two aspects; these correlations are not in agreement with the original conditions. It is caused by the fact, that the principal factor method selects an axis so as to maximize the total with the axis connected variance. We achieve then a solution of Fig. 5, in which the reference axes do not coincide with the  $x$ -vectors of Fig. 3; now a  $x$ -vector must be described by two reference axes.

This problem can be solved by a transformation of the reference frame; we know, that a rotation is always allowed. The problem suggests instantly the rotation by which the axes and the  $x$ -vectors coincide. By this rotation the correlations between aspects and the irrelevant  $x$ -variable disappear and an identification of the aspects has been achieved. We get a unique solution of which a causal interpretation meets with no difficulties. The result of this rotation is recorded in Table 6.



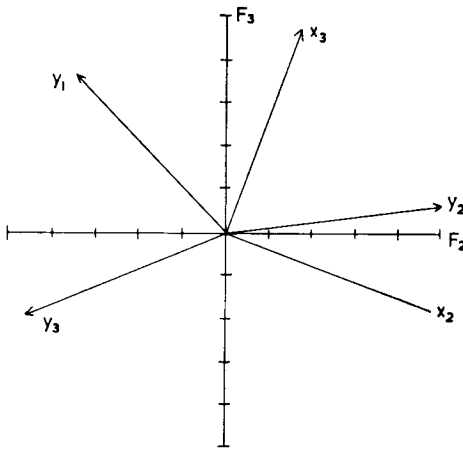


Fig.5 Positions of reference axes and vectors in the principal-factor solution (model with 2 independent variables).

Table 6 Caumax solution; model with 2 independent variables

Variable	Aspect		
	$F_1$	$F_2$	Sum $a_{im}^2$
$x_2$	1.00	0	1.00
$x_3$	0	1.00	1.00
$y_3$	-0.73	-0.68	1.00
$y_2$	0.89	0.46	1.00
$y_1$	-0.89	0.44	0.99

Table 7 Principle-factor solution; model with 4 independent variables

Variable	Aspect				
	$F_1$	$F_2$	$F_3$	$F_4$	Sum $a_{im}^2$
$x_1$	0.83	-0.39	-0.04	0.40	1.00
$x_2$	-0.54	-0.61	0.18	0.55	1.00
$x_3$	-0.07	0.62	-0.30	0.72	1.00
$x_4$	0.12	0.31	0.93	0.14	1.00
$y_3$	0.93	-0.23	0.29	0.02	1.00
$y_2$	-0.91	0.31	0.26	0.01	0.99
$y_1$	0.68	0.73	-0.04	0.01	1.00
Eigenvalue	3.15	1.68	1.15	1.01	6.99

A causal interpretation of the results obtained by this rotation can now be easily given, with the already mentioned limits however.

The significance of this rotation is much more apparent, if we have to do with more-dimensional systems. We shall demonstrate this by an example, in which the four independent variables  $x_1, x_2, x_3$  and  $x_4$  are involved and the reference frame is four-dimensional. The results of the aspect analysis are recorded in Table 7.

These results are rather disappointing. It is true, the number of aspects is equal to 4 according to the expectations — the number of independent variables was 4 — the causal background of the used equations, however, has mainly disappeared. It is difficult, even with our foreknowledge, to interpret some aspects. A relatively large proportion of the variance has been accounted for by the first aspect. Each aspect has relatively strong correlations with more than one  $x$ -variable. Identification and consequently the final solution are achieved again by a rotation by which the reference axes must coincide as well as possible with the  $x$ -vectors. In this case a solution is obtained in agreement with the original conditions. Table 8 gives the result of this rotation.

Table 8 *Caumax solution; model with 4 independent variables*

Variable	Aspect				Sum $a_{im}^2$
	$F_1$	$F_2$	$F_3$	$F_4$	
$x_1$	1.00	0	0	0	1.00
$x_2$	0	1.00	0	0	1.00
$x_3$	0	0	0	1.00	1.00
$x_4$	0	0	1.00	0	1.00
$y_3$	0.86	-0.30	0.32	-0.28	1.01
$y_2$	-0.88	0.36	0.23	0.19	0.99
$y_1$	0.29	-0.81	0.27	0.43	1.00

The usefulness of the rotation performed and the significance of the method of rotation are demonstrated clearly by this result. The rotation has achieved an identification of the aspects and the causal interpretation has been simplified.

The results obtained have demonstrated the importance of the following procedure. In order to obtain a causal interpretation the results of each factor analysis must be transformed by a rotation. The rotation criteria are the independent variables only, of which the number must be taken equal to the number of aspects found by the analysis. The rotation must try to get a situation in which every reference axis coincide as well as possible with one independent variable. The correlations between an aspect and the other independent variables must be made as small as possible. Finally, the rotation must be performed with all variables by means of the rotation matrix upon the  $x$ -variables. This rotation, which we have named *caumax-rotation*, aims to identify the aspects and to maximize the possibility towards a causal interpretation.

**Discussion**

We mentioned in the introductory section, that the factor analysis is used especially to formulate hypotheses about complex phenomena. In such an explorative investigation the researcher is still far away from a situation, in which he has a clear notion of the system of relationships and the estimation of the parameters of models as

Table 9 *Caumax solution; model with 4 correlated independent variables*

Variable	Aspect				
	$F_1$	$F_2$	$F_3$	$F_4$	Sum $a_{im}^2$
$x_1$	0.89	-0.03	0.45	-0.13	1.01
$x_2$	0.11	0.95	0.23	-0.20	1.00
$x_3$	-0.27	-0.29	-0.22	0.89	1.00
$x_4$	0.22	-0.07	0.97	-0.02	1.00
$y_3$	0.74	-0.10	0.59	-0.28	1.00
$y_2$	-0.90	0.32	-0.12	0.28	1.00
$y_1$	0.09	-0.85	0.11	0.51	1.01

described on p. 40 ff. are seen as very important. The researcher, employing the factor analysis, has mostly a more simple aim and tries:

- a) Firstly to examine how many independent variables do influence the system, and
- b) secondly to indicate which variables can be considered as independent ones in order to achieve an interpretation of the aspects.

The technique of the determination of the number of aspects or dimensions of the vector configuration will not be discussed, for this we refer to the literature. However, we have to realize that error factors in the equations may raise the number of dimensions. If these discrepancies are relatively unimportant, they can be neglected without any objection; the analysis produces in this case a number of eigenvectors with relatively large eigenvalues. If these errors are more important, the analysis will produce eigenvectors with relatively small nor large eigenvalues. Now, the researcher will be left in doubt of the real dimensions. He has to decide intuitively how large the number of dimensions is. An uncertainty as to the number of aspects exists in such cases. This uncertainty is enlarged by the presence of correlations between the independent variables; a correlation does have the tendency to decrease the dimensional number.

In connection with this we are interested in the answer to the question how strongly the results of a factor analysis are influenced by these disturbing factors and circumstances. In order to get an impression we have performed some computations on the influences of errors (standard deviation about 15% of the mean) and of the correlations between the independent variables. It appeared from these computations, that these disturbances in the used models had only little influence and that factor analyses are yielding rather consistent results. This can be illustrated by the result of an analysis (with rotation) on data obtained from the model of Fig. 1, in which the independent variables had rather strong correlations (0.20, -0.45, 0.64, -0.53, 0.20 and -0.27). The result is given in Table 9.

A comparison between these data and those of Table 8 demonstrates that the presence of these rather strong correlations between the independent variables did not change essentially the loadings and the character of the aspects<sup>3</sup>.

Concerning point b) we note, that the researcher has mostly to renounce of the objective and automatic rotation methods. The best known methods are those, that try to achieve the *simple structure* and in which a distinction between independent and

<sup>3</sup> An oblique rotation makes the correlations between aspect and irrelevant  $x$ -variables disappear.

dependent variables plays no role; all variables are used to find the new positions of the reference axes and the rotation itself can be performed without intervention of the researcher (Harman, 1960). Parenthetically, it will be clear that we reject the rationale of the simple structure approach.

On the other hand, the caumax rotation demands from the researcher a standpoint with regard to the most acceptable interpretation of the aspects, the independence of the variables to which the reference axes has to be rotated etc. An automatic solution is not obtained, if a marking of the variables with respect to their independence is difficult. In such cases the researcher, supported by his knowledge and the quantitative results of the analysis, will decide upon the final rotation only after an iterative rotation procedure. This also means, that the researcher achieves only during the investigation and rotation itself a logical interpretation of the aspects.

Summarizing, we have in such cases a marking of the variables with regard to their independence in the rear of the procedure. This rests on the reasonable but only empirically testable proposition, that the reverse of the proposition as described under the heading *Factor analysis and rotation* also holds: rotation to interpretable aspects leads to a rotation upon independent variables. In practice both procedures will be used at the same time, at which the accents will shift to the first or to the second way depending on the knowledge about the system.

From this the possibility also follows that different opinions on the character of the variables exist. In that case and that phase of the research the researcher must admit that more than one position of the reference frame leads to aspects that are interpretable.

Finally, we note that it is only allowed to make causal inferences from the counting rule of the number of independent variables, if all relevant variables are taken up in the analysis. Indeed, it is possible that an independent variable can not be measured or measurements are just missing. Such an independent variable does give a dimension to the vector configuration. After rotation to a certain position we have obtained an aspect too, of which the loadings can be considered to be caused by an absent variable. It is clear, that the counting rule holds only if this last variable is also counted.

## References

- Cattell, R. B., 1965. Factor analysis — An introduction to essentials. *Biometrics*, 21 : 190–215; 405–435.
- Ferrari, Th. J., 1963. Causal soil–plant relationships and path coefficients. *Plant Soil*, 19 : 81–96.
- Ferrari, Th. J., 1964. Auswertung biologischer Kettenprozesse mit Hilfe von Pfadkoeffizienten. *Biometr. Z.*, 6 : 80–102.
- Ferrari, Th. J., 1965. Models and their testing: considerations on the methodology of agricultural research. *Neth. J. Agric. Sci.*, 13 : 366–377.
- Ferrari, Th. J., Pijl, H. and Venekamp, J. T. N., 1957. Factor analysis in agricultural research. *Neth. J. Agric. Sci.*, 5 : 211–221.
- Harman, H. H., 1960. *Modern Factor Analysis*. Univ. Press, Chicago.
- Liberg, A. H. J. and Mol, J., 1961. Multiple factor analysis as a method of agricultural research. *Neth. J. Agric. Sci.*, 9 : 269–280.
- Mol, J., 1964. Modèles d'exploitations agricoles. C.E.E. Collection d'études, Ser. Agriculture No. 13, Bruxelles.
- Mol, J., 1966. De economische onderzoeker en zijn ruimte. *Economie*, 30 : 361–376.
- Seal, H., 1964. *Multivariate Statistical Analysis for Biologists*. Methuen, London.