

Multiple factor analysis as a method of agricultural research

A. H. J. LIBERG and J. MOL

Agricultural Economics Research Institute, The Hague, Holland

1. Introduction

If you see a playground during recreation, swarming with children it seems quite a chaos to you. After a while, however, you will discover some pattern in the moving crowd. Some children are playing with each other, some are quarreling, others are behaving themselves independent of what happens around. This situation is a characteristic feature of economic life. First of all there is a crowd of variables. In farming you have to deal with many input factors and many output factors. There is the structure of the farm (arable plan and different kinds of animals), the efficiency in the use of labour and machineries, the productivity of the soil and the animals etc. Second you will see that many of these variables change simultaneously. So an extension of the cattleherd may be related to a change in the use of labour and in the winning of home-grown feedingstuffs. If you compare two farms you will see that they differ not only in one respect, but in many.

What is the pattern underlying these differences? Are they ruled by certain laws? Just as the swarming children on the playground some variables are playing with each other and are positively correlated, some are quarreling and are negatively correlated, others are behaving themselves independently from each other. Just as a child, a variable can at the same time play with the first one, quarrel with the second and be indifferent for the behaviour of the third one. What are the rules of the games and of the behaviour of the children? What are the rules controlling the relation between our variables, how should we describe economic life in a simple model?

The economist, analysing farm records, is in the same position as the observer of the playground wanting to know why children play in this or that way. There is a lot of empirical material and the difficulty is how to interpret it. The problem of the economist is to describe an economic chaos in a non-chaotic economy. It has been our experience that the multiple factor analysis is a useful tool in helping to solve the problems of the economist. Before applying this method to a farm economic problem we will illustrate some characteristic features of the multiple factor analysis by a theoretical example.

2. Characteristic features of multiple factor analysis

2.1. Vectorial representation of original data

We want to illustrate the features of the multiple factor analysis with a simple example, dealing with few firms and few variables. We will assume that we know

Received for publication 30th August, 1961.

the following variables of a number of N firms, delivering a certain product to the population centre X :

O = yearly production.

C = value of the stock of capital goods.

D = distance to the centre.

A = age of the firms.

M = degree of mechanisation, measured by the amount of capital per worker.

W = yearly wage of a worker.

P = average price received during the year.

These variables are measured as deviations from their averages; TABLE 1 contains the variables for some firms.

TABLE 1. Original data of some firms

Variables	Firms					
	I	II	III	IV	N-I	N
O	-0,8	+3,2	+2,4	—	—	—
C	-0,4	+1,6	+1,2	—	—	—
D	+1,0	-4,0	-3,0	—	—	—
A	-0,7	+2,9	+2,2	—	—	—
M	+3,0	+0,75	0,0	—	—	—
W	-0,5	+2,0	+1,5	—	—	—
P	-6,0	-1,5	0,0	—	—	—

The table is artificially constructed so that some variables are positively correlated, some negatively and others are independent from each other. A positive correlation exists between O and C (the value of O is $2 \times$ the value of C). There is a negative correlation between O and D, (the value of O $- 0,8 \times$ the value of D). Further we notice that there is no correlation between M on the one side and O, C, W, A and D on the other side; so is for instance row M multiplied by row D: $(+3,0) \times (+1,0) + (+0,75) \times (-4,0) + (0,0) \times (-3,0) = 0$. It is supposed that these relations between the rows exist also for the parts of rows not filled up. The readers, acquainted with matrix algebra will see that the rank of the matrix is 2.

The information, given in the table can also be shown in a 3-dimensional picture. The 7 rows can be represented as 7 vectors. FIG. 1 gives the position of vector D in the 3-dimensional space.

The ordinates of D are respectively: +1, -4 and -3. If you draw also the other 6 vectors you will see that vectors O, C, A and W are in the opposite direction of vector O; vectors M and P become situated in the plane through axes I and II and are perpendicular to vector D and the others (see FIG. 2).

In analogy with this figure the variables of all N firms can be thought as vectors in a N dimensional space. For this N dimensional space the same terms as for the 3-dimensional can be used. So we can say that in the N -dimensional space the vectors O and C are in direct line with each other and that vector M is to vector D perpendicular.

We see that for our firms O and C are closely related (for each firm $O = 2 \times C$); this can be expressed in terms of geometry as: "O and C are pointing in the same direction."

In this case there is a complete positive correlation between variable O and vari-

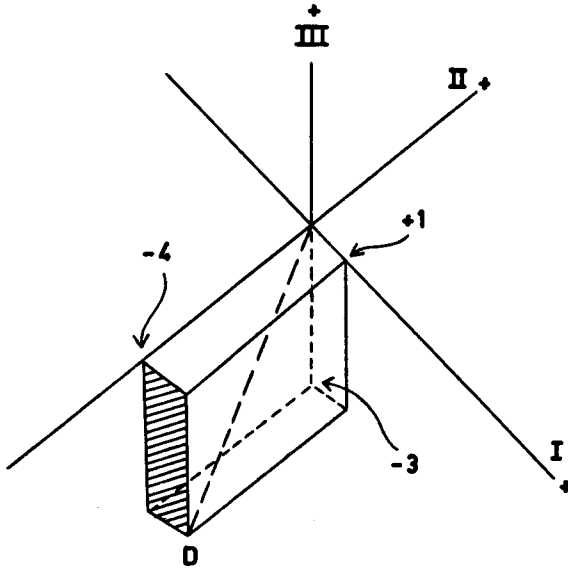


FIG. 1. Position of the vector D.

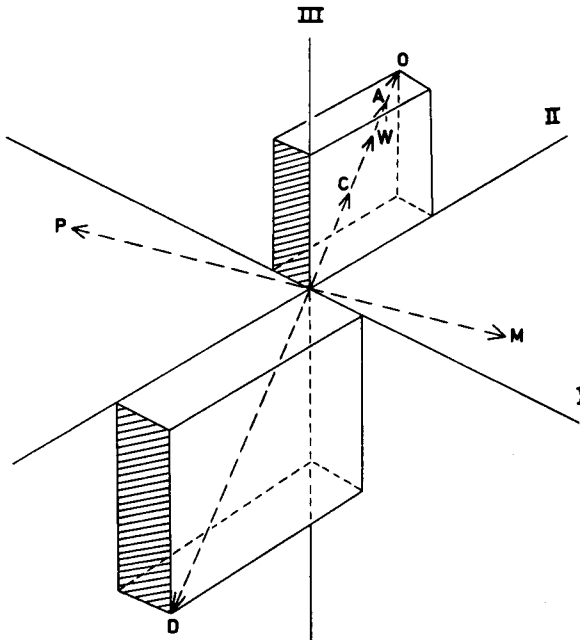


FIG. 2. Position of the vectors O, A, W, C, D, P and M.

able C (+1) and a complete negative correlation between variable O and variable D (—1). This can be illustrated for instance by a graph where the ordinates represent respectively variable O and variable C; the dots will be on one straight line. The correlations between variable M and the variables O, C, D, A and W are zero. The sum of the products of row M with one of the others is zero.¹ It will be seen that, where the vectors form an angle of 90°, the correlation is zero and when they form an angle of 180° the correlation is —1; the correlation coefficient can thus be measured by the size of the angle between the vectors. An angle between 0° and 180° means a correlation coefficient between 1 and —1.

As can be seen from FIG. 2 the whole configuration of vectors is contained in a 2-dimensional space.² This is often considered as a characteristic figure of multiple factor analysis viz. the representing of the basic traits of a large number of chaotic data which always can be represented in a space of high dimension in a space of minimum number of dimensions.³ In doing this, the attempt is made to eliminate the stochastic deviations. If for instance M was situated at a very short distance of the plane in which the other vectors are contained and you look upon this as a stochastic deviation, then you still consider the configuration of the vectors as being contained in a two-dimensional space (If in this space M should be projected, this vector M should become shorter as now in FIG. 2).

The position of the vectors in the 2-dimensional space is illustrated in FIG. 3.

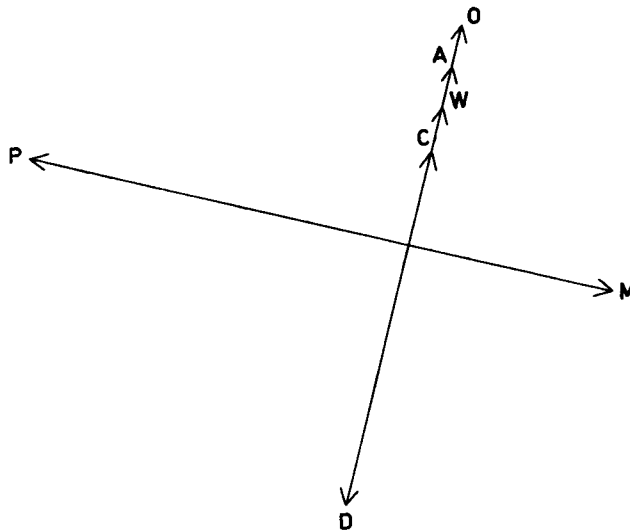


FIG. 3. Position of the vectors O, A, W, C, D, P and M in a plane.

¹ The correlation coefficient is: $r = \frac{\sum x_1 x_2}{N \sigma_{x_1} \sigma_{x_2}}$ (x_1 and x_2 are deviations from their average). If $\sum x_1 x_2 = 0$ then $r = 0$.

² We recall that the rank of the table is 2.

³ HOLZINGER, K. J. and HARMAN, H. H.: Factor analysis, a synthesis of factorial methods. Univ. Chicago Press 1941, p. 3—6.

2.2. Formulating of hypotheses

For the interpretation of the configuration of vectors shown in FIG. 3 different hypotheses can be made. Suppose we should not have measured the variables M and P and that we had only to interpret the configuration of the vectors O, A, W, C and D, then different hypothesis are consistent with the data. Let us give two theories.

1. The size of the firm, measured as the yearly production (O) is larger in the direction of the centre, this negative correlation between O and D is caused by the higher wages (W) in the centre. High wages stimulate substitution of labour for capital and this sort of investment in general stimulates large scale production.

2. It is not necessary that the size of the firm is influenced by the high wages.

The relation between wages and the costs of mechanisation may be just so advantageous for a high degree of mechanisation for firms near the centre as well as for firms at a large distance from the centre. The reason for differences in size may be very simple. The firms in the centre are old ones and the firms outside the centre are younger ones. Usually young firms are smaller than old ones.

Both hypotheses give a reasonable explanation of the relations between the variables.

Often it is possible that by taking in account more variables one of the hypotheses can be rejected as improbable. Also this is a characteristic feature of the multiple factor analysis: many variables are considered simultaneously.

Doing this and looking now at the variable M we will see that hypothesis number 1 is improbable because the degree of mechanisation (M) has nothing to do with the variables D, C, A, W and O (vector M is perpendicular to vectors D, C, W, A and O). There is only a strong negative correlation between M and the price received for the product (P). On basis of this relation the hypothesis can be stated that heavy mechanized firms make a product of a lower quality and therefore make a lower price. This price has nothing to do with the distance to the centre (D).

2.3. Deviations from the example

In practical applications of the multiple factor analysis the vectors are not always as beautiful as shown in FIG. 3 in perpendicular directions. Often you will find configurations as shown in FIG. 4.

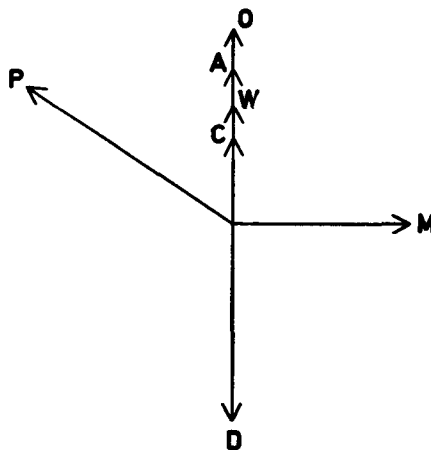


FIGURE 4

This configuration says that the price of the product (P) has a negative correlation with the degree of mechanization (M) as well as with the distance to the centre (D). The differences in prices (P) have now not only to do with the mechanization-quality-aspect, but also we must consider (D) the aspect of the distance (transportation costs) as another explanation for the differences in prices. A schematic presentation of the multiple factor analysis could in this case be as is shown in TABLE 2.

TABLE 2. Table of aspects¹

Variables	Distance-age aspect	Mechanization-quality aspect
O	+	0
C	+	0
D	—	0
A	+	0
M	0	+
W	+	0
P	+	—

In this table the first column shows that the decrease of the distance to the center (D = —) is coupled to the growing older of the firms (A = +), an increase of the size of the firm (O = + and C = +). The firms near the centre receive a higher price (P = +) and this has probably to do with transportation costs. The degree of mechanization has nothing to do with this aspect (M = 0).

The second aspect shows another cause of price difference. High mechanized firms (M = +) receive for their products a low price (P = —).

In the multiple factor analysis not only the patterns of variable behaviour are given by + and — signs, but also quantitative propositions are made about them. In short we shall say something about this.

The oscillation of a variable may be measured in different ways. A well known measure is the standard deviation (σ). An other measure is the variance, the square of the standard deviation (σ^2). The multiple factor analysis indicates the degree in which the variance of a variable is connected to a certain aspect. In a case as shown in FIG. 4 we get propositions as for instance; the variance of P is for x % connected with the aspect of the distance (aspect 1) and for y % with the mechanization-quality aspect (aspect 2). These percentages can easily be read off the graphs if the vectors in the N-dimensional spaces all are given the same length (for instance 10 cm); this is possible by measuring every variable in a certain scale.

By multiplying every number of row one of TABLE 1 by $\frac{1}{\sigma \sqrt{N}}$, (σ being the standard deviation of variable O, vector O would become a length of 10 units. Applying this for every row and representing these "normalized" data again graphically, the above pictured configurations of vectors would only alter in this way that the vectors all would get the same length; the angles between the different vectors would

¹ The quadrant to the right of A and above M is called the positive quadrant in FIG. 4 and 5. If we had taken another quadrant as positive, f.i. to the right of D and below M than in this table + signs in the first aspect would have become — signs and — signs would have become + signs. This is not essential; the interpretation of the distant-age aspect would not have altered.

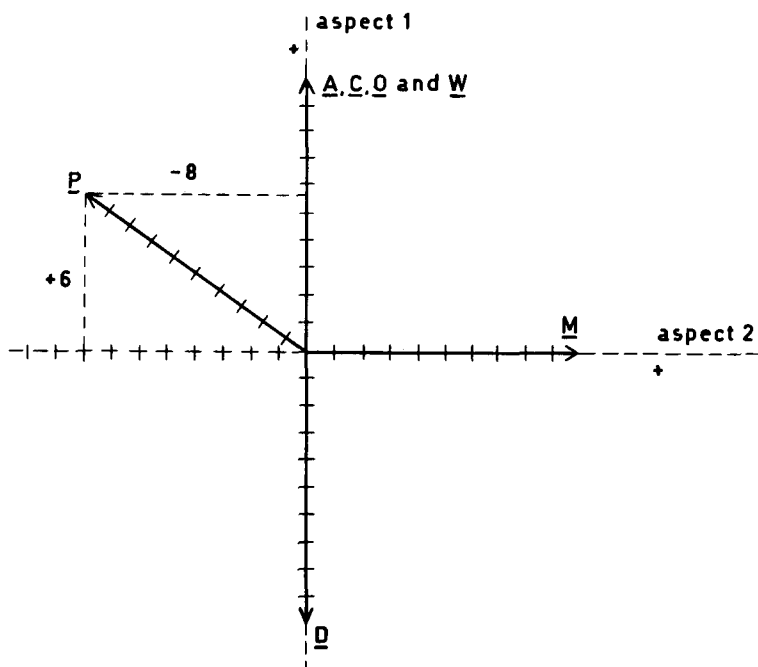


FIGURE 5

stay unaltered. We suppose that in FIG. 5 has been worked with this "normalization" (and that the vectors here also are contained in a space of just two dimensions). Without explanation we state that the configuration of FIG. 5 leads to this proposition: "the variance of P is connected with the second aspect for $(8)^2 = 64\%$ of its variance and connected with the first aspect for $(6)^2$ or 36% of its variance". Are you inclined to see D and M as "independent" variables and P as dependent on them, then your proposition can be "the distance to the centre explains 36% of the price-variance, the degree of mechanization 64% ". The table of aspects, quantified and in accordance with FIG. 5 is shown in TABLE 3. In our highly hypothetical example the total variance of the variables can fully be described by the two aspects mentioned. In applied research, however, often only

TABLE 3. Table of aspects

Variables	Percentages of variance connected with:		Communality
	distance-age aspect	mechanization-quality aspect	
O	100 +	0	100
C	100 +	0	100
D	100 -	0	100
A	100 +	0	100
M	0	100 +	100
W	100 +	0	100
P	36 +	64 -	100

a part of the total variance can be explained because of stochastic deviations, non-linear connections between the variables, not taking up important explanatory variables in the analysis; at the other side variables may be put into analysis, which have only minor relations with other variables. Such a variable is for instance variable Q, represented in FIG. 6 below; the angles between vector Q and the other vectors are all $\pm 90^\circ$ and this means almost a zero-correlation between Q and the other variables.

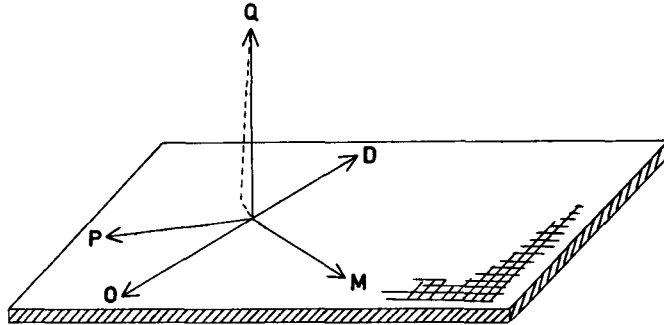


FIGURE 6

Variable Q thus is a "lonely" variable; the "uniqueness" of Q is very high. Variable Q is unimportant for explaining the pattern of behaviour of the variables D, M, O, and P. A projection of the above configuration in the plane (through D, M, O and P) gives us the graph of FIG. 7.

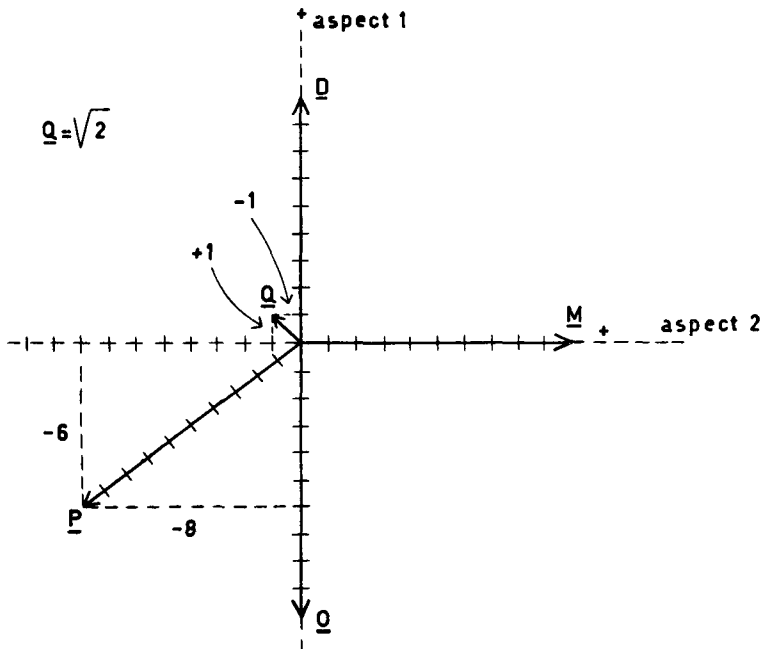


FIGURE 7

GRAPH 7 reveals that only $(Q)^2 = \sqrt{2^2} = 2\%$ of the total variance of Q can be connected to the remainder of variables. The table of aspects (TABLE 4) is now in accordance with FIG. 7.

TABLE 4. Table of aspects

Variables	Percentage of variance connected with:		Communality
	aspect 1	aspect 2	
D	81 +	0	81
O	100 +	0	100
M	0	100 +	100
Q	1 +	1 —	2
P	36 —	64 —	100

Most of the time we will find in applied research more than two aspects; this makes the identifying of aspects often more difficult. We shall not draw further on this subject in this article.

3. An application of multiple factor analysis

After the explanation of some characteristic features of the multiple factor analysis we now give the results of a research of HAMMING and LIBERG, who applied the multiple factor analysis on data of a sample of 200 mixed farms on sandy soil in the Netherlands.¹

The matrix of empirical data was of the nature that 7 aspects had to be distinguished. The table of aspects (TABLE 5) gives the percentages of the variances connected with the different aspects.

A work unit is the standardized labour requirement of a crop or an animal per year. The total work units of the farm may be considered as a measure of the size of the farm. The different enterprises are added according to uniform labour requirements for all farms. The number of total farm-work units per man expresses in a certain way the efficiency of the use of labour; the labour income and machinery costs per farm-work unit expresses the remuneration for labour and capital left by the productivity of the soil and the animals or in other words left by the productivity of the different work units.

The structure of the farm has been measured by the percentages of the pig work units, cash-crop work units etc. of the total farm-work units (Variables 4, 5, 6, 7 and 8). If the farms obtain an average productivity of the enterprises the contribution of the different enterprises to the remuneration of labour and capital (variables 15, 16, 17 and 18) must harmonise with the farm structure.

The aspects I and II describe two different independent causes for the labour income per man. The first cause is the efficiency of labour, expressed by the number of work units per man. In the first column we see that an increasing of the variable "number of work units per man" is coupled to the growing of the labour income

¹ HAMMING, G., and A. H. J. LIBERG: "Aspecten van de bedrijfsvoering van gemengde bedrijven op zandgrond". Landbouw-Economisch Instituut 1960.

TABLE 5. Table of aspects of mixed farms

Variables	Percentage of variance connected with aspects							Communitarity
	I	II	III	IV	V	VI	VII	
<i>Income, labour efficiency and productivity</i>								
1. Labour income per man	55+	45+	.	3+	.	.	.	103 ²
2. Total farm-work units per man	81+	.	.	.	3—	.	.	84
3. Labour income and machinery costs per farm-work unit	79+	.	.	4+	.	.	83
<i>Structure of the farm</i>								
4. Cash crop work units as a percentage of total farm-work units	.	.	92+	92
5. Pig work units as a percentage of total farm-work units	85+	.	.	.	85
6. Poultry work units as a percentage of total farm-work units	12—	10+	62+	.	.	84
7. Horticultural work units as a percentage of total farm-work units	4+	4+	.	19+	.	27
8. Cattle work units as a percentage of total farm-work units	42—	38—	12—	4—	.	96
<i>Size of the farm and labour supply</i>								
9. Total farm-work units	10+	.	13+	.	5—	.	58+	86
10. Acreage cultivated land per farm	3+	.	16+	6—	10—	10—	42+	87
11. Total family wages per farm	25—	.	6+	.	3—	.	18+	52
12. Family wages per ha	45—	.	14—	6+	15+	6+	.	86
13. Paid wages as a percentage of total wages	7+	.	3+	3—	.	3—	9—	52
14. Total farm-work units per ha	.	.	23—	32+	15+	14+	.	84
<i>Profitability of the different enterprises</i>								
15. Return of cash crops per farm-work unit	92+	92
16. Returns minus feeding costs of pigs per farm-work unit	5+	.	74+	.	.	.	79
17. Returns minus feeding costs of poultry per farm-work unit	6+	7—	6+	62+	.	.	81
18. Returns minus feeding costs of cattle per farm-work unit	8+	40—	28—	15—	3—	.	94

¹ The percentages 1, 2 and 3 are replaced by dots.

² Due to roundings off, this percentage has increased above 100 %.

per man. This can be obtained without altering the structure of the farm because the variables 4, 5, 6, 7 and 8 have practically nothing to do with aspect I. How is this number of work units obtained? The main factor is to look for low costs of family wages per ha. This regards 45 % of the variance in the family wage per ha. There are 2 ways in obtaining low family wages per ha: viz. caring for few family members working on the farm or getting more acreage. Aspect I shows that the variance in the acreage of the farm is related to this aspect for only 3 % and the size of the working family for 25 %. Besides this there is also an indication that

farms with somewhat more paid labour obtain also a good labour efficiency. In general the mixed farms studied are family farms with only some seasonal paid labour. So there may be a big chance that an over supply of family labour exists. Paid labour then is an indication for a strong shortage of family labour on the farm. It is allowed to change all signs of a column. In that case we read the relations in the contradictory way. If there is a lot of family labour on the farm the labour efficiency as well as the income per man will be low.

The correlation coefficient between the acreage of the farm and the labour efficiency is 0,31. The fact that in aspect I the acreage of the farm has only a slight relation with the labour efficiency may be due to the fact that the farms studied are a sample out of efficient farms without a clear oversupply of family labour.

In another publication of the Agricultural Economics Research Institute¹ this aspect has been worked out in a sample out of all mixed farms. As in this sample also the inefficient ones and the very small ones are included (one ha and up), the influence of the acreage is higher. In a similar multiple factor analysis as given in TABLE 5 it has been shown that 19 % of the variance in acreage was related with the same aspect I.

In the sample of 2900 farms, aspect 1 is illustrated by TABLE 6.

TABLE 6. Labour efficiency on 2900 farms; number of work units per man

Type of labour supply	Farm acreage (ha)					All farms
	1-7	7-10	10-15	15-20	> 20	
Farmer alone	2550	3300	3750	4800	—	3100
Farmer with paid labour	3050	2950	3350	3800	4250	3650
Farmer with son(s)	1950	2250	2700	3150	3550	2550
Farmer with other family members	1900	2350	3000	3300	4150	2650
All types	2300	2750	3000	3500	4000	2900

In this sample labour efficiency increases by increasing acreage and by the type of labour supply. Clearly it is here also shown that the best efficiency of labour is obtained by the farmer alone (low family wages) and by the farmer with paid labour.

Another remarkable point already indicated is that the structure of the farm has no relation to the efficiency in the use of labour. Theoretically it is possible that the labour efficiency is raised by attracting more labour intensive enterprises as poultry- and pigholding. The empiric data show that the farmers do not solve their labourproblems by intensifying the structure of the farm.

Further the structure of the farm has no correlation with the productivity of the different enterprises as shown in aspect II. A higher income per man will be obtained if the contribution of the enterprises to the remuneration of labour and capital is higher (variables 16, 17 and 18) as is expected on basis of the farm structure (variables 4, 5, 6, 7 and 8). This aspect shows that farmers will not specialize in that enterprise where they can make the highest profit. This is also shown by the aspects III, IV, V and VI. In these aspects the labour income per man, the labour efficiency and the profitability of the enterprises do not play any role.

¹ MARIS, A., and R. RIJNEVELD: "Het kleine-boerenvraagstuk op de zandgronden". Landbouw-Economisch Instituut 1960.

The aspects III, IV, V and VI deal with some interesting things on the structure of the farms. These aspects are related with the acreage of the farms and with the intensity of farming (number of work units per farm). Large farms are specialized on cash crops (aspect III), smaller farms on pig-holding (aspect IV), the smallest farms on poultry and some types of horticulture (aspects V and VI). If we look to the changes from large to small farms we see that the increase in family wages per ha parallels the increment in work units per ha. The consequence is that there are no differences in labour efficiency (work units per man) between large farms and small farms.

On the basis of the aspects I through VI the conclusion may be drawn that there is a good adaptation between the acreage of the farm and the farm structure, but no adaptation between the supply of family labour and the farm structure. As will also be clear from TABLE 6 the farm structure and the intensity of farming do not react on changes in labour supply due to the family cycle. The young farmer has a good labour efficiency. If he gets older and his son, who will succeed, grows up, the work to be done remains the same. So in that phase of the family cycle the labour efficiency is low.

Aspect VII indicates that large farms have more work units and also higher family wages or in other words small farms have less labour units and also less family labour. This aspect indicates that there is much similarity in the kind of farming of small and large farms. The large farms in comparison with small farms do not use to a high extend their natural opportunity in obtaining a better labour efficiency. This may be a serious problem. If a small farmer has only one son at his farm it will be quite normal. But, as aspect VII indicates, if a large farmer has e.g. 2 sons on his farm there may rise difficulties in succeeding the father. As in Holland farms are very scarce to buy, it may be that too much labour stays in agriculture on larger farms.

We hope in the above example of a multiple factor analysis to have proved that this method of analysis is an important tool for empiric economic research. In a simple table of aspects many variables are considered in their relations with each other. The different independent tendencies controlling farming may be clearly seen.

It will be clear that the multiple factor analysis is only the starting point of a research project. After this analysis, it is necessary to interpret the found connections and tendencies with concrete figures, as is done in TABLE 6. After that it is necessary to confront the empirical reality with the theoretical possibilities to solve the problems indicated.

In the case above, the problems which call for a solution are e.g. how to improve the labour efficiency, how to diminish the family labour supply, especially on the larger farms, how to employ these labour forces outside the agriculture and what measures are necessary to make the farmers adapt the intensity of farming to the changing labour supply due to the family cycle.