FACTOR ANALYSIS, A FARM MANAGEMENT RESEARCH TOOL

R. C. Duncan*

Factor analysis, a statistical technique used for many years in the fields of psychology and education, is described. Limitations of its use are discussed and these are illustrated with reference to some recent applications of the technique in farm management research.

1 INTRODUCTION

The primary motivation for this examination of factor analysis was the prospect of its usefulness in analysing farm survey data. It has been presented in the literature as a technique which could aid interpretation of the relationships between a large number of interdependent variables. An assessment of the technique seems desirable, in view of the anticipated increase in its use in farm management research. Evidently, some of those prompted to use it have not been fully aware of its limitations. Furthermore, confusion has arisen because of the existence of various schools of thought on the subject.

In this article, a brief description of factor analysis is followed by an examination of its claimed applicability. The various avenues of application and their limitations are examined with particular reference to three examples of the use of the technique in farm management research.

2 FACTOR ANALYSIS—A STATISTICAL TECHNIQUE

At this point the following definition is appropriately non-committal:

> Factor analysis supplies methods for reducing a large number of observed variables to a lesser number of in some way more fundamental variables or, as they are usually called, factors. This is usually done through an analysis of intercorrelations between the observed variables.[1]

---

* Economics Research Officer, Grafton Agricultural Research Station.

The author wishes to thank John L. Dillon for his comments on earlier drafts, with the usual caveat about responsibility for errors and omissions.

[1] S. Henrysson, *Applicability of Factor Analysis in the Behavioral Sciences, A Methodological Study* (Stockholm: Almqvist & Wiksell, 1957), p. 14.

Factor analysis is one of a number of statistical techniques which comprise the branch of statistical theory known as multivariate analysis. Multivariate analysis is concerned with the relationship of sets of dependent variables, e.g., a set of N individuals each of whom exhibits values for n different variables giving a total of Nn observations. Note that "the (n) variates are dependent among themselves so that we cannot split off one or more from the others and consider it by itself".[2]

According to Kendall, multivariate analysis can be divided into two parts "according to whether we are concerned with *dependence* or *interdependence*".[3] In the analysis of dependence "we are interested in how a certain specified group (the dependent variates) depend on the others". In the analysis of interdependence "we are interested in how a group of variates are related among themselves, no one being marked out by the conditions of the problem as of greater prior importance than the others".[4] Factor analysis is of the latter type of multivariate analysis.

In essence, the problem dealt with by factor analysis is as follows. Given Nn observations on n variables, can we find a set of h new (hypothetical) variables, preferably with $h < n$, which will account for all, or nearly all, of the variation in the original Nn observations. Thus the chief goal of factor analysis is sometimes described as the attainment of parsimony of description, wherein a solution will yield new variables (factors) which convey all the information contained in the original set of variables.

The question arises as to what, precisely, does the factor solution add to the information already held. Can the factors be interpreted as "fundamental" or "underlying" variables which determine the structure of the field of study? Most of the controversy over the use of factor analysis stems from attempts to interpret and/or identify the new derived variables.

THE FACTORIAL SOLUTION

Various techniques have been devised to obtain the factor solution. These differ in method and in the solution derived, but basically they fall into two distinct groups known as Component Analysis and Factor Analysis. The laborious detail of the different computing procedures will not be discussed here.[5]

---

[2] M. G. Kendall, *A Course in Multivariate Analysis* (London: Charles Griffin, 1961), p. 5.

[3] *Ibid.*, p. 6.

[4] M. G. Kendall, "Factor Analysis as a Statistical Technique", *Journal of the Royal Statistical Society*, Series B, Volume 12, Number 1, (1950), pp. 60–73.

[5] For full details of the different computing procedures—see H. H. Harman, *Modern Factor Analysis* (Chicago: University of Chicago Press, 1960).

Hotelling[6] has developed what is known variously as the Principal-Components, Principal-Factor or Principal-Axis solution. In Hotelling's solution *components* or *factors* are chosen so that the first factor explains as much as possible of the total variance in all variables. The second factor is chosen so as to be uncorrelated with the first, and to explain as much as possible of the residual total variance, and so on.

The number of factors extracted by the Principal-Factor solution will be less than n only if some of the variables are linearly dependent on another. However, a reduction in the dimensions of variance can be approximated because the first few factors (which account for as much of the variation as possible in descending order) may account for say 85 to 90 per cent of total variance. "We can then say that the variation is represented *approximately* by the first two or three variates and in favourable circumstances may be able to neglect the remainder."[7] In comparison with other solutions such as the Centroid solution[8], Hotelling's solution (and its lineal descendants) is to be preferred on statistical grounds, being a least squares solution. Calculation of the Principal-Factor solution is extremely laborious however, and has been made computationally feasible only with the advent of computers.[9]

3 HYPOTHESIS TESTING OR HYPOTHESIS GENERATING

The main developments in factor analysis are due to psychologists, and the language of the technique reflects its non-statistical origin. The technique as a whole[10] has not gained wide acceptance by statisticians, nor are many statisticians familiar with it. Psychologists have been using and developing factor analysis since the turn of the century. Much of the application of factor analysis in the field of psychology could be summed up as follows:

> Many psychologists have engaged in extensive testing programs, employing factor analysis to determine a relatively small number of tests to describe the human mind as completely as possible. The usual approach includes the factor analysis of a large battery of tests in order to identify a few common factors. Then the tests which best measure these factors, or, preferably, revised tests based on these, may be selected as direct measures of the "factors of mind".[11]

---

[6] H. Hotelling, "Analysis of a Complex of Statistical Variables into Principal Components", *Journal of Educational Psychology*, Volume 24, (1933), pp. 417–41. 498–520.

[7] Kendall, A Course in Multivariate Analysis, *op. cit.*, p. 11.

[8] See Harman, *op. cit.*, Ch. 10.

[9] For the computing procedure see W. W. Cooley and P. R. Lohnes, *Multivariate Procedures for the Behavioral Sciences* (New York: John Wiley, 1962).

[10] It is probably safe to say that statisticians, if concerned with this section of multivariate analysis at all, prefer to separate Component Analysis and Factor Analysis and to regard the latter technique dubiously.

[11] Harman, *op. cit.*, p. 6.

The important questions which arise in regard to the application of the technique are (i) whether factor analysis can be used as an exploratory tool in an "unstructured" field to *generate* hypotheses; (ii) whether it can be used only for *testing* hypotheses; or (iii) whether it can be used for both purposes.

The case for the use of factor analysis in generating hypotheses has been put by Cattell.[12]

> Factor analysis provides also a method far more free than most methods from the necessity to elaborate rigid hypotheses. It is the ideal method of ***open exploration*** in regions unstructured by present knowledge.[13]
>
> Particularly in the biological and social sciences the researcher is presented with so bewildering a multitude of possible variables that unless he first factorizes to find the inherent organization or *structure*, i.e., to find which surface variables are representatives of more significant, less numerous underlying variables, an immense waste of effort could (and does!) take place.[14]
>
> Of course the factorist enters with *some* hypothesis even when he seems to enter with none. He enters an experiment with the hypothesis ***that some structure exists to be discovered.***[15]

Cattell is concerned with two points. Firstly, that "it *is* possible to observe covariation and to develop laws *without* theories".[16] Secondly, by being wholly concerned with the rejection of a hypothesis the researcher may fail to observe covariation which could lead to an improved hypothesis. "The essential thing to observe is any and every evidence of *law*, i.e., of orderly covariation, in the field concerned."[17]

The case for the use of factor analysis only as a tool for testing hypotheses is supported by Burt, who has said that:

> Factor analysis is required where experimental analysis (at any rate for the time being) is impracticable. Properly applied, it should . . . always be regarded as "confirmatory", even where it *appears* to be "exploratory". It is not a means of "discovering" factors hitherto unsuspected.[18]

This implies that there is an illogical step in the reasoning supporting the case for factor analysis as a hypothesis-generating technique. The domain or field of study must be defined before the observations are carried out. The only way in which the field of study can be defined is by *a priori* assumptions about the causal relationships. In effect, the researcher *must* be testing some hypothesis.

---

[12] R. B. Cattell, *Factor Analysis* (New York: Harper, 1952).

[13] *Ibid.*, p. 14.

[14] *Ibid.*, p. 16.

[15] *Ibid.*, p. 21.

[16] *Ibid.*, p. 13.

[17] *Ibid.*, p. 13.

[18] C. Burt, "Discussion: Factor Analysis", *Journal of the Royal Statistical Society*, Series B, Volume 12, Number 1, (1950), p. 87.

However, Babington-Smith[19] argues in support of factor analysis as an exploratory technique. He maintains that if factor analysis is only confirmatory, "then I am inclined to maintain we are on a circular track, that the structure of the tests is constrained by our preconceptions, so that the results only confirm whether we have tests which fit our preconceptions".

Henrysson[20] holds that it is difficult ". . . to maintain any clear distinction between testing and generating of hypotheses. By hypotheses-testing some of the hypotheses can be disproved, and at the same time a basis can be found for establishing of new ones". It is quite reasonable to suggest that ideas for a new hypothesis can arise from the testing and rejection of a hypothesis. It is maintained here however, that the role of hypothesis-generating is of a lower order.

What is implied in the confirmation of a hypothesis by factor analysis? If certain causal relationships are postulated, it is hypothesized that certain variables will be highly correlated. By the nature of the factor solution it is possible to see if certain variables group together in the same dimension. It should be possible therefore to test several hypotheses (or one hypothesis composed of several distinct parts) in a single factor analysis of a body of data.

The rationale behind Cattell's advocacy of factor analysis as an exploratory tool differs from this interpretation as follows. Where it is found that several variables have high loadings (correlation coefficients) on one factor it is assumed that these variables are in fact different manifestations of the same variable, and therefore redundancy exists. Further, on a more complex plane, by observing which variables have high loadings on the factors and which variables do not, one is able to "interpret" the underlying structure revealed.

A warning about the interpretation of observed correlation in general, has been given by Peak[21]:

> . . . an obtained correlation between two tests simply demonstrates that some functional unity exists, within the limits set by the size of the coefficient and by the ambiguities of its interpretation. From this information alone it is impossible to say what the common aspect is and whether it is simple or complex.

---

[19] B. Babington-Smith, "Discussion: Factor Analysis", *ibid.*, p. 93.

[20] Henrysson, *op. cit.*, p. 91.

[21] H. Peak, "Problems of Objective Observation", Ch. 6 of L. Festinger and D. Katz, *eds, Research Methods in the Behavioral Sciences,* (New York: Holt, Rinehart and Winston, 1953), pp. 273–4.

A similar warning has been given by Barnard[22] in a specific reference to factor analysis.

> . . . factor analysis attempts to isolate *qualities* out of quantities—to isolate, for example, the quality of left-handedness out of the quantitative measures. This method is bound to confuse distinct qualities which have quantitative relations with each other. Boyle's law, for example, says that pressure is porportional to density. For a given mass of gas, therefore, the correlation coefficient between pressure and density is 1. This would lead a factor analyst to say that pressure is the same thing as density.

In summary, then, we should be aware that when covariation is observed between two variables, one of three interpretations may be made:

(*a*) that different manifestations of the same variable have been observed;

(*b*) that both variables bear some functional relationship to each other or to some other variable; or

(*c*) that the covariation is due to accident.

Barnard is talking about the situation which applies in (b). The school of thought represented by Cattell assumes that only type (a) correlation is present in the observed data; moreover, even if the researcher admits that type (b) is also present, it is contended that when both forms are present it is impossible to distinguish between them.

It seems to be widely agreed that the use of factor analysis as a *descriptive condensing aid*, i.e., to reduce redundancy, cannot be criticised. However, it is contended that the position is not so clear-cut. The validity of the use of factor analysis in this way depends on the use which is made of the factor solution, and the choice of data.

One of two different situations is implied. Firstly, if it is *known* that all the variables measure the same variable, then the factor analysis is redundant. If it is only *thought* that they measure the same variable, then a hypothesis is being tested and should be stated explicitly beforehand. Correlation of types (b) and (c) above cannot be ignored. Secondly, where the researcher is hoping to reduce the dimensions of variance present in a large set of variables, the analysis may reveal that most of the total variance is found in the first two or three dimensions. In this case it is important to know what use is to be made of the solution. Without restrictions as to the data contained in the observation matrix, it seems that two valid uses exist. Firstly, the factors could be thought of merely as reference axes, and the observed variables described in terms of these reference points (by factor loadings). Or, the reference axes could be used in some way as artificial variables to describe the dimensions present in the set. Secondly, by deleting those variables with high loadings on unimportant factors, the observation matrix could be "purified" in some sense. It is difficult to envisage much scope for these applications of the technique, as compared with its use in the testing of hypotheses.

[22] G. H. Barnard, "Discussion: Factor Analysis", *Journal of the Royal Statistical Society*, Series B, Volume 12, Number 1, (1950), p. 91.

4 ON THE USE OF FACTOR ANALYSIS IN FARM MANAGEMENT RESEARCH

A study of personal and social factors affecting, or thought to affect, successful communication to farmers of advice by extension services, was made by Brien, Wrigley and Jardine.[23] Thirteen personal and social variables were measured. These variables were measured by scales constructed from selected questions. Factor analysis was used to reduce the set of thirteen variables to a smaller set of more meaningful, "more nearly uncorrelated, derived scales".[24] The resulting set of six derived scales was correlated with two different measures of farmer performance.

Obviously the authors intended to use factor analysis as a condensing aid. They state:

> . . . unlike the items of a scale which all measure the same or closely similar variables, the scales themselves are intended to measure different variables. However, when as many as thirteen scales are constructed it is not immediately evident to what extent they are really different. Factor analysis offers a reasonably objective way of examining this matter and, in the present case, suggests that six scales, if properly constructed, contain most of the information of the original thirteen scales.[25]

The authors were endeavouring to eliminate redundancy amongst the variables. Only if it is possible to dismiss the likelihood of correlation due to the existence of functional relationships between the observed variables is it valid to say that only redundant variables would be eliminated. Some of the original scales considered were: conceptual skill, urbanization, situational motivation and attitude to printed matter. It would be a difficult enough matter to define these variables, let alone to say whether or not there was any functional relationship between them.

The type of factor analysis which can be carried out on variables such as were included in the above study is of the hypothesis-testing type, illustrated here in a study reported by Hobbs.[26] The project "was designed to measure the value orientation of farm operators along with some other personal and social characteristics and relate these to the economic performance of the farms managed by these individuals."[27]

---

[23] J. P. Brien, J. F. Wrigley, and R. Jardine, "A Study of Some Personal and Social Factors in Relation to Farmer Performance", this *Review*, Volume 33, No. 3 (September, 1965), pp. 126–46.

[24] *Ibid.*, pp. 133–4.

[25] *Ibid.*, p. 135.

[26] D. Hobbs, "Use of Factor Analysis in a Farm Management Study", Paper presented at Symposium on *Present Use and Potential of Linear Programming and other Operations Research Techniques in Farm Management Extension*, (University of Missouri, Columbia, Missouri, Jan., 1965).

[27] *Ibid.*, p. 9.

The basic behavioural model from which the hypothesis to be tested was generated was formulated as follows:

> Managerial behaviour was considered to be a function of (a) predispositional forces such as beliefs, values and attitudes (b) abilities, skills and capacities of the individual and (c) the social, psychological, economic situation in which the individual acts. These factors are considered to be interdependent but each is also considered to contribute some unique or independent variance in predicting and accounting for human behaviour. . . . From a factor analysis viewpoint these three dimensions may be considered as hypothesized factors among the dependent variables.[28]

Twelve variables were hypothesized as measures of the three conceptual dimensions postulated above. Factor analysis was used to provide statistical support for the hypothesized behavioural model, i.e., to see if the selected variables designed to measure the different dimensions would group together in different factors in the factor solution.

Hobbs goes on to look at some possible applications of factor analysis in farm management. He suggests studies (similar to that reported in his paper) relating farm operator goals, values, attitudes, and biographical data to various criteria of farm operator performance, together with an analysis of the relationship between adoption of farm practices and farm performance.

Hobbs also suggests that factor analysis could be useful in studies of farm records by "gaining further insights concerning the underlying functions of all of the various performance indicators included in farm record summaries".[29] He suggests the large number of performance indicators could be organized into a few hypothesized dimensions. For instance, the hypothesis that a number of indicators measure, say, efficiency in different ways, could be tested. These indicators could be combined in some sort of overall index of efficiency and thus possibly lead to a reduction in the number of indicators of concern to the farm advisor. A hypothesis such as this could be tested under different farming conditions.

Choosing between the large number of possible farm performance indicators is a problem. However, it is doubtful whether Hobbs' approach to the problem would be an efficient method of attack. Perhaps a better approach would be through the estimation of functional relationships suggested by biological and economic theory.

Workers at the Agricultural Economics Research Institute, The Hague, Netherlands, claim to have developed factor analysis as a tool of farm management research, in particular as a comparative research tool for

---

[28] *Ibid.*, p. 10.

[29] *Ibid.*, pp. 17–8.

analysis of survey data.[30] Their investigation relates to 100 and 160 dairy farms respectively in two different areas of Friesland Province. The choice of variables included in the factor analysis was "partly based on previous knowledge and on assumptions about the causal relationships. But the ultimate conclusion about the relationships between the variables [was] derived from the study of the structures exposed by factor analysis".[31]

The authors chose the variables on the basis of some hypothesis (not stated) about the causal relationships. However, they use the technique as one of open exploration, "as a method by which the data are used to get the most sensible interpretation of the correlations present".[32] The use of a "sensible interpretation" (rationalization?) of observed correlations seems to be a far cry from a scientific approach to research.

As well as "interpreting" and naming the factors, the authors attempt, from observation of variables covarying in each dimension, to make recommendations for more efficient organization of the individual farm. It is in this way that they conceive of factor analysis as a comparative analysis tool. Analysis of this kind appears to be stretching the interpretation of observed correlation far beyond breaking-point.

It will not be denied that, given a proper use of factor analysis, some worthwhile observations can be made regarding improvements in individual farm organization. However, a point which should be borne in mind is that the above study is an analysis of the differences between farms, not of the make-up of individual farms.

5 CONCLUSIONS

An attempt has been made in this paper to delineate the limitations of the factor analysis technique. Its principal application was seen to be the testing of hypotheses; its use as a condensing aid to reduce redundancy must be controlled by stringent restrictions on the nature of the variables under study.

With reference to farm management research in particular it is apparent that factor analysis is a technique without any great degree of finesse. Still it does appear to have a place in the analysis of survey data. During the course of a survey a number of tentative hypotheses are frequently formulated, quite apart from initial hypotheses around which the survey

---

[30] J. de Veer, "Farm Management Research on Dairy Farms in the Netherlands, with Special Reference to Factor Analysis, and the Need for Co-operation with Technical Scientists", *O.E.C.D. Documentation in Agriculture and Food*, No. 71, pp. 127–60.

[31] *Ibid.*, p. 132.

[32] *Ibid.*, p. 133.

may have been designed. A common procedure in analysing survey data in the past has been to tabulate a great number of variables and present them with some subjective assessment as to the way in which the results confirm or deny the hypotheses under consideration. This is no more than a procedure in which one does intuitively what factor analysis does objectively. Factor analysis could be useful, therefore, in preliminary testing of hypotheses, especially where a large number of variables are concerned. Further, there seems to be no reason why more than one hypothesis cannot be tested in the same analysis or more than one procedure of testing used.