THE ECONOMIC
AND SOCIAL
RESEARCH INSTITUTE
MEMORANDUM SERIES
NO. 104

FACTOR FICTION

A Monte Carlo Approach to Factor Analysis

by

Brendan J. Whelan

(Seminar Paper delivered at ESRI on 17 June 1975)

September 1975

Factor Fiction - A Monte Carlo Approach to Factor Analysis

1. Introduction

Factor analysis is a complex technique. The derivation of even its most basic results requires an understanding of matrix algebra, and the mathematics involved in establishing the more complex rotational methods is beyond the algebraic competence of the vast majority of researchers. Indeed in some cases factor analytic techniques are used whose distributional properties are not yet known, or are known only for impracticably large samples. Asymptotic results are of little value unless one can be sure that they will not be seriously in error for samples of the size usually encountered in practice.

This complexity gives rise to several problems. In the first place, it is difficult to teach factor analysis, especially to those with a limited mathematical background. I speak from bitter experience on this subject since all my attempts to teach factor analysis seem to end up as either incomprehensible algebra, or vague and unsatisfactory verbal descriptions. Secondly, users of the technique are forced to rely on hunch and rule-of-thumb in the absence of clear-cut distributional results for "small" (non-asymptotic) samples. This difficulty is further aggravated by the very wide variety of modifications and refinements which have developed in the literature and whose strengths and weaknesses are not well known.

The present paper suggests a partial solution to these problems. The basic idea is to generate, by computer, a sample from a population with a known factor structure.* Various factor analytic techniques can be applied to this sample and the estimates so derived may then be compared with the known factor structure. This short paper cannot pretend to be an exhaustive treatment of the application of Monte Carlo methods to factor

---

* An appropriate program is presented in the Appendix.

analysis. The best I can hope to do is to indicate the broad areas to which this technique may be applied, and to present some initial, illustrative results.

Many applications can be envisaged for the Monte Carlo approach. First of all, students with little knowledge of mathematics often find a carefully constructed and well explained example more palatable than reams of algebra. Stuart (1962) shows how an arithmetic example may be most effectively used to teach the basic principles of sampling theory. The first part of the present paper attempts to explain factor analysis in a similar way. By setting up a population where the relationships between the factors and the variables are known, it is hoped to make students conscious of the hypothetical model underlying all applications of factor analysis and to show how this model may be estimated by means of the various factor analytic techniques. Interested students could generate samples from different factor structures by means of the program presented in the Appendix, and could then apply the various programs to estimating these structures.

A second possible application of the Monte Carlo approach to factor analysis is the testing of the accuracy and efficiency of various programs. The results of such tests on three currently available programs are presented in the second part of this paper.

Thirdly, Monte Carlo methods may be used to shed some light on the distributional properties of factors. Lawley and Maxwell (1963) have made considerable progress from an algebraic point of view on these topics, but many important results have not yet been derived. Large scale Monte Carlo experiments could be the answer to these difficulties. However, such experiments would have needed more time (both computer time and researcher time!) than I had at my disposal and I do not deal any further with distributional properties in this paper.

The fourth area where Monte Carlo methods could be of use (and here we come to the title of the paper) is in investigating the effects of various practical limitations which are met in almost all applications. For example one is never sure how many factors to extract. Rules of thumb abound but their validity is rarely checked, and the costs (errors) involved in extracting the wrong number of factors are unknown. Another limitation which traditional theory does not take account of is the fact that researchers frequently wish to include discontinuous variables in a factor analysis. The implications of using such variables as opposed to continuous, normally distributed variables are not well known: Monte Carlo methods could help clarify this issue.

The third part of this paper is therefore devoted to a consideration of some of these issues for one particular factor structure. I hope to illustrate the extent to which commonly used factor analytic procedures can unearth the "true" underlying dimensions of a set of variables - in short the extent to which factor analysis produces fact or fiction.

2. Factor Analysis explained by example

Let us assume that we are conducting an attitudinal study and have administered a questionnaire containing 16 variables (items or sets of items) to 200 subjects. The scores for the first 10 subjects are shown in Table 1. Let us further assume that these 16 variables reflect only four underlying independent* factors. If, say, we were investigating the general area of intelligence, the original items might be individual intelligence tests of various sorts (either individual items or averages over sets of items) and the underlying factors might be, say, verbal ability, quantitative ability, three dimensional visualization and creativity. Saying that the original items reflect only these four factors means that if one could know an individual's

* For the sake of simplicity, I deal only with uncorrelated factors in this paper. In practice, the factors may be correlated (oblique) and certain techniques (PROMAX, OBLIMIN) can be used to estimate them.

score on each of the four factors, one could calculate his score on any of the 16 variables in the questionnaire. In other words, all the information we have about an individual is contained in his scores on the factors. The problem in practice is that we cannot observe factor scores; all we can observe are test scores. Yet it would be great interest to try to estimate these factor scores since they provide a much simpler and more parsimonious description of the data. This is what factor analysis tries to do.

Just for the purposes of illustration let us assume that an omniscient being (OB) has revealed the relationships between factor scores and test scores to us for a certain set of data. These relationships are known if we know the co-efficients a_{j1}, a_{j2}, a_{j3}, a_{j4}, in the following equation for variable j (which we call z_j)

$$z_j = a_{j1} F_1 + a_{j2} F_2 + a_{j3} F_3 + a_{j4} F_4$$

There will be 16 such equations, one for each variable, all with different co-efficients. Table 2 shows the full set of such co-efficients, as revealed by the OB, for our example. Thus, to calculate an individual's score on variable one from his factor scores we would multiply his score on Factor 1 by -0.083, his score on Factor 2 by 0.714, his score on Factor 3 by 0.621 and his score on Factor 4 by -0.313 and add the results, i.e.

$$(-0.083) \cdot (F_1) + (0.714) \cdot (F_2) + (0.621) \cdot (F_3) + (-0.313) \cdot (F_4)$$

If our obliging OB can tell us an individual's factor scores, then we can use Table 2 to predict (with certainty) what that individual's score on any of the 16 variables will be.

Apart from their use in predicting the variables from the factors, the co-efficients or 'loadings' in Table 2 give the correlations between the variables and the factors. Thus, the correlation between z_1 and F_1 is 0.083

and that between z_1 and F_4 is -0.313. Squaring the co-efficient a_{ij} gives the proportion of the variance in variable i attributable to changes in factor j. For instance, factor 2 accounts for 50.9 per cent ($= 0.713^2$) of the variance in variable 1.

It should also be noted that we have assumed the factors to be independent (uncorrelated). This implies that any significant correlation which we observe between two variables is due to the fact that they are both related to the same factors. A fundamental factor theorem states that the correlation between variables 1 and 2 (r_{12}) may be decomposed as follows

$$r_{12} = r_{1F_1} \; r_{2F_1} + r_{1F_2} \; r_{2F_2} + r_{1F_3} \; r_{2F_3} + \ldots$$

where r_{kF_m} is the correlation of variable k with factor m i.e. the loading of factor m on variable k. In terms of our example, the correlation between variable 1 and variable 2 should, from the theorem, be equal to

$$r_{12} = 0.002 + 0.660 - 0.089 - 0.110$$
$$= 0.463$$

The observed correlation from the correlation matrix given below (Table 5) is 0.405

Table 2 allows us to calculate the variables from the factors. Sometimes it is useful to do the opposite and calculate the factors from the variables. To see how this might be done, consider the meaning of the first four rows of Table 2 i.e.

$$z_1 = 0.083\ F_1 + 0.714\ F_2 + 0.621\ F_3 + 0.313\ F_4$$

$$z_2 = 0.018\ F_1 + 0.925\ F_2 - 0.144\ F_3 + 0.350\ F_4$$

$$z_3 = 0.233\ F_1 + 0.952\ F_2 - 0.017\ F_3 - 0.199\ F_4$$

$$z_4 = -0.287\ F_1 + 0.032\ F_2 + 0.947\ F_3 + 0.139\ F_4$$

These are four equations representing the variables expressed as weighted sums of the factors. By successively eliminating, say, F_2, F_3 and F_4 from equation 1, we can obtain Factor 1 (F_1) expressed as a function of z_1, z_2, z_3 and z_4. F_2, F_3 and F_4 can be similarly expressed. The result will be the following set of equations:

$$F_1 = 14.390\ z_1 + 2.066 z_2 - 12.487\ z_3 - 9.344 z_4$$

$$F_2 = -2.020\ z_1 + 0.112 z_2 + 2.410 z_3 + 1.384 z_4$$

$$F_3 = 5.436\ z_1 + 1.046 z_2 - 5.012 z_3 - 2.439 z_4$$

$$F_4 = 6.813 z_1 + 2.880 z_2 - 7.770 z_3 - 4.168 z_4$$

These equations are based on the first four rows in Table 1. Selecting any four rows (variables) will allow any factor to be expressed in terms of these variables.

Essentially, we have now completed a factor analysis of the data; we can express any variable as a weighted sum of the factors, and any factor as a weighted sum of the variables. In terms of our example, we can say exactly how each test score is based on the factor scores, and how the factor scores may be derived from the test scores.

However, our example is seriously unrealistic from two points of view. (i) It assumes that any variation in a variable is attributable only to variations in the factor scores. That is, there is no source of variation in the data other than the four factors.

(ii) It assumes that the variables can be measured without error.

Both these over-simplifications may be overcome if we modify the original equation linking variables and factors to read

$$z_i = a_{i1} F_1 + a_{i2} F_2 + a_{i3} F_3 + a_{i4} F_4 + d_i U_i$$

where U_i is a random variable which includes all sources of variation unique to variable i and d_i is a co-efficient. U_i has a mean of zero and variance of 1, so that the mean of $d_i U_i$ is zero and its variance d_i^2. The variance of z_i (= 1, since $z_i \frown N(0, 1)$) may thus be partitioned in two parts, that attributable to fluctuations in the scores on the common factors (known as the communality) and that attributable to the unique component (known as the uniqueness). The communality for variable i is given by $\sum_{j=1}^{4} a_{ij}^2$ and the uniqueness is given by d_i^2.

$$\text{Var}(z_i) = 1 = \sum_{j=1}^{4} a_{ij}^2 + d_i^2$$

If we know the communality h_i^2 (= $\sum_{j=1}^{4} a_{ij}^2$) we can calculate d_i^2 by $1 - h_i^2$.

A knowledge of the relationship between 16 variables and four factors will now entail knowledge of 80 co-efficients (64 loadings a_{ij} and 16 unique co-efficients d_i). This is illustrated in Table 3 which shows the scores of 10"individuals" as generated from the factor pattern (set of co-efficients) shown in Table 4. This pattern includes the full eighty co-efficients.

Having examined the complications which arise when the unique term is included, let us now return to the factor pattern shown in Table 2 and consider how we might estimate this set of co-efficients if our OB was on a day off. The only information now at our disposal is the set of test scores as shown in Table 1. From these, we can calculate the correlation matrix (Table 5). We saw above that these correlations arise because the variables are determined by the factors. Examination of the correlation matrix can therefore shed some light on the nature of the factor structure.*

Submitting this correlation matrix to a factor analysis program will produce estimates of the co-efficients a_{ij} which express the variables as functions of the factors (the factor pattern matrix). It will also produce estimates of the proportion of the total variation in the set of variables attributable to each of the factors. Table 6 shows the factor pattern and proportions of variance as estimated by the SPSS program, together with the "true" co-efficients used to generate the data. The agreement between the estimates and the true values can be seen to be very good, since the correlation between the true and estimated co-efficients ranges from 0.997 for factor 4 to 0.991 for factor 3.

However, and here we come to a vital and sometimes misunderstood point, the co-efficients as shown in Table 6 are not the only ones which could account for the observed correlation matrix. Indeed, there is an infinity of such sets, and, mathematically, there is no way of choosing between them. One alternative set is shown in Table 7. To verify that these co-efficients can account for the observed correlations just as well as those in Table 6, let us re-examine the decomposition of r_{12} which we carried out above (p. 4). The fundamental factor theorem states that -

---

* Cluster analysis proceeds by grouping the items in the correlation matrix and has been shown by Raven [1971] to produce similar results to factor analysis. See also McQuitty [1957].

$$r_{12} = r_{1F_i} \; r_{2F_1} + r_{1F_2} \; r_{2F_2} + r_{1F_2} \; r_{2F_3} + r_{1F_4} \; r_{2F_4}$$

Using the co-efficients as estimated in Table 6, we obtain $r_{12} = 0.006 + 0.560 - 0.043 - 0.118 = 0.405$. Using those from Table 7 gives

$r_{12} = -0.026 + 0.645 - 0.028 - 0.185 = 0.405$

Thus, each set of co-efficients is equally effective in explaining the observed correlations. An investigator who has no information other than the correlation matrix must therefore devise new criteria on the basis of which to choose between the alternative sets of co-efficients. Several such criteria have been suggested, the most famous of which is probably Thurstone's concept of "simple structure" and its modifications [see Thurstone 1947]. Essentially, this involved selecting the set of co-efficients which was most easily interpretable in terms of subject matter of the inquiry. "Interpretability" is judged on the basis of the number of high (near 1.0) and low (near zero) loadings. Various criteria can be used depending on whether one wants to simplify the rows, the columns or both the rows and the columns of the factor matrix. These are known as "rotational" criteria and familiar types are QUARTIMAX and VARIMAX.

To sum up the discussion so far, I would like to draw attention to three aspects of factor analysis which I believe are frequently neglected. The first of these is the fact that we are operating with a <u>linear</u> model. As Morrison (1967) points out "the (linear) model of factor analysis is as much part of our hypothesis about the dependence structure as the choice of exactly m common factors ..... If the covariances reproduced by the m-factor linear model fail to fit the sample values adequately,

it is as proper to reject linearity as it is to advance the more usual finding that m common factors are inadequate for explaining the sample correlations." While a full non-linear treatment may be "frightfully complex" (Harman), researchers should, I think, reflect on the substantive theory of the topic in question to make sure that a linear model is really relevant. Appropriate transformations of the variables before analysis, as is frequently done in regression, might help solve this problem.

The second point is that the number of factors extracted may determine the answers obtained. The implications of extracting the wrong number of factors is further explored in section 4 below.

The third aspect which I would like to emphasise is that the choice of a certain rotational method (e.g. VARIMAX) implies certain hypotheses about the relationships between the factors and the variables. Specifically, such a choice defines one set of co-efficients as better than the alternative (mathematically equivalent) sets on the basis of a certain definition of "simplicity" or "interpretability". Researchers should be clear that, in using a particular rotational criterion, they are opting for one of several possible definitions of interpretability, and that this definition will affect the substantive nature of the results obtained.

3. Accuracy and Efficiency of Three Programs

The three programs discussed in this section are:

(1) the Statistical Package for the Social Sciences (SPSS) from the University of Chicago

(2) PCVARIM, a program belonging to E. E. Davis and based on the Cooley - Londes tri-diagonalization procedures

(3) the IBM scientific subroutines package (SSP).

Several sets of data generated by the program FACGEN given in the Appendix were submitted to each of these programs. These were all small data sets (200 observations, 16 variables, 4 factors) but they should indicate the relative merits of the programs.

Table 8 shows the Central Processing Unit (C.P.U.) time taken by each of the programs to analyse the two data sets. In each case this meant extraction of 6 down to 2 factors with varimax rotation of each set of factors, the data being input from tape. It may be seen that PCVARIM is considerably more efficient than the other programs. It should be noted that this comparison is somewhat unfair to the SSP program, since it is assumed that the time taken by the SSP program to extract and rotate 6 down to 2 factors is 5 times that taken to extract and rotate 4 factors. This is probably an over-estimate since it assumes re-computation of the correlation matrix at each stage.

The amount of information printed out by the various programs varies somewhat. All three can print out the means, standard deviations, correlation matrix, the varimax-rotated factor pattern matrix, the eigenvalues, communality estimates and estimates of the proportion of explained variance to total variance. The SPSS program can produce, in addition,

the inverse and determinant of the correlation matrix, the factor matrix rotated by the QUARTIMAX, EQUIMAX or OBLIQUE criteria, the factor score co-efficient matrix and a graphical plot of the rotated factors.

All three programs will accept the correlation matrix as input, and, as Bent et al (1970) point out, this procedure can save the user "enormous" amounts of machine time. Experiments inputting the present 16 variable correlation matrix showed that the SPSS program could carry out a factor analysis in about 20 seconds CPU time. Inputting the correlation matrix is therefore to be recommended.

The accuracy of three programs is assessed in Table 9. It may seem that all three programs produce practically perfect estimates of the parameters of the structure with no uniqueness. The estimates for the parameters of the structure with a unique component are not so close, but the average correlation of over 0.88 for each program would seem quite satisfactory.

The conclusion of this section is that PCVARIM is more efficient than the other programs. However, SPSS does have the range of additional options outlined above, and on occasion these may be useful.

4. Two Common Problems in the Use of Factor Analysis

(i) How many factors exist in the data?

In the case of a factor structure with no unique variance, such as that shown in Table 2, this question can be unambiguously solved by inspection of the eigenvalues (and consequently the percentage explained variance). To verify this consider the eigenvalues derived from the correlation matrix shown in Table 5 (which was based on a factor structure with no uniqueness).

Factor	Eigenvalue	Percentage Variance	
1	5.74	35.9	} = 100%
2	4.65	29.0	
3	3.78	23.6	
4	1.83	11.4	
5-16	0.0	0.0	

Each eigenvalue after the fourth is zero. We can therefore be certain that the structure can be completely expressed in terms of 4 factors.

Problems arise when one attempts to apply the same logic to a correlation matrix which was generated by a structure containing unique variance. Consider the eigenvalues of the correlation matrix generated by the structure shown in Table 4 which does have a unique component.

Factor	Eigenvalue	Percentage Variance
1	3.34	20.9
2	3.05	19.1
3	2.42	15.1
4	1.74	10.9
5	0.70	4.4
6	0.66	4.1
7	0.64	4.0
8	0.52	3.3
9	0.49	3.1
10	0.44	2.7
11	0.42	2.6
12	0.40	2.5
13	0.36	2.2
14	0.29	1.8
15	0.27	1.7
16	0.26	1.6

The first factor accounts for 20.9 per cent of the variance, the first two for 40.9 per cent and so on. When should one stop extracting factors, and assume that the unexplained variance is simply due to the influence of the random errors summarized in the unique component? At least three ways of solving this question have been suggested. Firstly, one can extract as many factors as there are eigenvalues greater than 1.0. The rationale for this rule of thumb is that the total variance in the data set is equal to 16 (since each of the 16 variables is standardized to unit variance) and hence those factors with an eigenvalue greater than one are explaining more than the "average" amount of variance.

Secondly, one can use a test suggested by Bartlett (1954). He showed that

$$\log \left\{ \left(\lambda_{k+1} \quad \lambda_{k+2} \quad \ldots\ldots \lambda_p \right)^{-1} \left\{ \frac{\lambda_{k+1} + \lambda_{k+2} + \ldots\ldots \lambda_p}{p-k} \right\}^{p-k} \right\}$$

multiplied by a certain multiplier is distributed as X^2. (p is total number of variables and k is number of factors which we think are adequate to explain the variation in the data. λ_j is the j-th eigenvalue). A significant value of the criterion suggests that an insufficient number of factors has been extracted.

A third possible approach, which may be combined with the other two, is to examine the patterns of (rotated) loadings generated by different numbers of factors and choose the most easily interpretable solution. This has the great advantage of ensuring that one's results make substantive sense, but it may involve a certain amount of subjectivity.

Let us now see whether these procedures give the "correct" answer when applied to our data set which we know contains four factors. As may be seen from the eigenvalues listed above, the "eigenvalues greater than one" rule gives the correct answer i.e. extract four factors. Bartlett's test for three to seven factors is shown below

Number of Factors Extracted	Number Remaining	Criterion	D.F.	Significance Level
3	13	360.5	90	$P < 99.5\%$
4	12	124.9	77	$P < 99.5\%$
5	11	101.3	65	$P < 99.5\%$
6	10	77.4	54	$99\% < P < 97.5\%$
7	9	50.5	44	$90\% < P < 75\%$

The test seems misleading in the present case. It suggests that there is a less than 0.5% chance that the last twelve factors arose from a complex of uncorrelated variables. Rigorous adherence to this test in a practical situation would lead one to extract six and not four factors.

It is not really possible in the present case to judge the usefulness of the third criterion, interpretability, because the loadings which were used to generate the factor structure were arbitrary and did not have any "simple structure" properties.* In a factor analysis involving real data one could invoke considerations of meaningfulness and parsimony to select the appropriate number of factors.

What happens if one extracts the "wrong" number of factors? Table 10 shows the three, four and five factor solutions for the data set generated by the co-efficients in Table 4. All three solutions show quite a marked similarity.

---

* Of course, there is no reason a "simple" factor structure could not be input into the FACGEN Program and an interpretable data set generated. It might even be possible to modify the program to create "simple" structures.

(ii) What happens when categorial data are used?

Quite frequently in economic and social research important variables can only be measured at a dichotomous or polychotomous level. The inclusion of such variables violates the assumption of normally distributed, continuous variables on which factor analysis is based. A related, though less serious, difficulty arises when variables are measured on a scale containing only three, five or seven points. The larger the number of points on such scale the more closely the variables conform to the assumption of continuity. Any number of points greater than four is generally assumed to provide a sufficiently close approximation to continuity for the purposes of factor analysis.

To investigate this problem in its most acute form, it was decided to dichotimize a set of data from the four factor structure shown in Table 2. This was done by generating the data set by the FACGEN program (see Appendix) and then transforming the resulting values of the variables by setting each negative value equal to $-\frac{1}{2}$ and each positive value equal to $+\frac{1}{2}$. The resulting data set therefore consisted entirely of $-\frac{1}{2}$'s and $+\frac{1}{2}$'s. This data set was factor-analysed by means of the SPSS program (principal factor method PA2) and the results, together with the "original co-efficients" used to generate the data, are shown in the first two sections of Table 11.

It may be seen that the actual and estimated structures are quite similar, the highest correlation observed between estimates and actual being over 0.9. It should, of course, be borne in mind that one would rarely in practice run a factor analysis on a correlation matrix derived entirely from dichotomous data. However, the factor pattern on which the present data set was based contained no unique component and when a more precise level of measurement was used correlations between actual and estimated factors were always over 0.99. The relatively small deterioration

in the quality of the estimates obtained when dichotomized data are used is therefore quite striking.

The last set of figures in Table 11 is an extension of the above test suggested to me by a discussion which I had with John Raven. He had had the experience with several "real" data sets that items with similar proportions of endorsement tended to be grouped together by factor analysis. That is, items with which a high proportion of the sample agreed tended to load on one factor, irrespective of their substantive content, while items with which a low proportion agreed tended to load on another factor.

The recurrence of this phenomenon had led John to suspect that this was an artifact of the factor analysis algorithm. Such "artificial" factors, if they exist, are clearly a nuisance when one is trying to make sense of one's results.

Varying rates of endorsement were simulated in the following way. A set of 200 observations on 16 variables, again based on the factor pattern shown in Table 2, was generated. The observations on variables 1-4 were dichotomized by allocating negative values to one category and positive values to the other. (Since all the variables are, before dichotomization, $\sim$ N (0, 1) this corresponded to assuming a 50 per cent endorsement rate). Variables 5-8 were dichotomized by allocating values below 0.68 to one category and those above it to the other (0.68 corresponds to the 75th parcentile of a unit normal variate. Therefore this corresponds to a 25 per cent endorsement rate). Variables 9-12 were similarly treated so that they had a 10 per cent endorsement rate and variables 13-16 a 5 per cent endorsement rate.

While the first four variables do load most heavily on the first factor, there seems to be no distinct pattern in the loadings of the other three sets of variables. Thus John's suspicion finds no confirmation in these results.

However, we see that even in the case of unequal endorsement rates, the estimates of the factors derived by the program are still quite good, since all the correlation co-efficients are greater than 0.75. Clearly, it would be of interest to see if this result holds good when a unique component is present.

References

Bartlett (1954) "Multiplying factors for various X^2 approximations" Journal of the Royal Statistical Society Series B Vol. 16.

Bent, D. et al (1970) "Statistics Package for the Social Sciences (SPSS) Manual".

Harman, H. (1967) "Modern Factor Analysis"

Lawley and Maxwell (1963) "Factor Analysis as a Statistical Method"

McQuitty (1957) "Elementary Linkage Analysis" Psychological Measurement, No. 17 p. 207.

Morrison (1967) "Multivariate Statistical Methods".

Raven, J. Ritchie, J. and Baxter, D (1971) "Factor Analysis and Cluster Analysis" Economic and Social Review, April 1971, Vol. 2 No. 3.

Stuart, Alan (1962) "Basic Ideas of Scientific Sampling".

Thurstone L. (1947) "Multiple Factor Analysis".

Table 1: Scores of 10 "respondents"on 16 variables.
No measurement or other errors.

Respondent No.																
1	1.8802843	1.5543957	1.8755312	0.5634809	1.8895292	-1.7137918	1.3797140	2.1312809	-0.9886680	0.3937969	1.4799786	-0.7318919	-1.3529720	0.8308043	-1.2952824	-1.0890036
2	-0.2237200	0.2471080	-0.6844826	0.6414865	0.2225040	-0.1348723	0.8212793	-0.2208400	-0.0390895	-1.6451302	-0.8723996	-1.8873672	0.0390476	-0.4576845	-0.6970932	-0.7984824
3	-0.2877959	-2.0585432	-1.0383272	0.4343902	-1.0001011	1.0083103	0.8334945	-0.2349048	-1.0719624	0.7332998	0.9118591	0.3413723	1.6361456	1.3900070	-0.9402488	1.5841455
4	2.3864565	1.8815784	2.5001459	0.7124277	2.3639059	-0.8226079	0.2963609	1.2903767	0.4493510	0.5708054	0.8316155	0.8234603	-2.2642870	-0.2865424	0.2119897	-0.1086026
5	-1.3853807	-0.7032669	-1.4063320	-0.2961699	-1.1389179	0.8575528	-0.7092476	-1.3855715	0.6579337	-0.8962489	-1.3418541	-0.3081454	0.7962834	-0.7892872	0.6476588	0.1245812
6	-1.5936842	-0.3722053	-0.7335233	-1.6276388	-1.5441141	-0.7991400	-0.7396988	0.2915913	-0.7742781	0.7723325	0.3437638	-0.0711684	1.1251278	0.7092866	-0.0384324	-1.0245447
7	0.9071494	1.1621389	1.0732479	0.1411002	1.0673103	-1.1055555	0.5356496	1.1283035	-0.3432116	0.0336391	0.5588196	-0.5189485	-0.9272390	0.1788452	-0.5105533	-0.9385645
8	0.3039175	-1.5842743	-1.3009424	1.7323112	-0.0365569	0.3687289	2.8071871	0.3897915	-1.8449697	-1.0293989	0.7206523	-2.3215504	1.4618187	1.5457544	-2.7033033	0.5590973
9	0.0132244	-0.0096582	-0.1971591	0.3070697	0.1069394	0.1963647	0.1738320	-0.2435173	0.1639695	-0.4953615	-0.3491162	-0.3588212	-0.0360738	-0.2554787	-0.0536493	0.0296811
10	-0.2014241	-0.0462776	-0.6914772	0.7972777	0.1550723	1.3701925	-0.4705115	-1.7012882	1.5907450	-1.6251030	-1.8943768	-0.1932245	-0.3743438	-1.7128258	1.0286770	0.6607755

Table 2: Co-efficients to estimate the variables from the factors (factor pattern matrix)

Variable	Factor			
	1	2	3	4
1	0.0832195	0.7136059	0.6211664	-0.3130381
2	0.0183298	0.9254011	-0.1436086	0.3502465
3	0.2339566	0.9515763	-0.0166898	-0.1987171
4	-0.2874494	0.0319199	0.9471081	-0.1390697
5	-0.0519799	0.8221784	0.5656589	-0.0367599
6	-0.7611248	-0.4298670	0.3215278	-0.3640375
7	0.4660682	-0.1779892	0.8661066	0.0309699
8	0.9083121	0.4170263	-0.0153999	0.0286597
9	-0.9303820	0.3524770	-0.1006791	0.0037600
10	0.6035870	0.1443493	-0.5643373	-0.5443974
11	0.9217330	0.1044291	-0.0171199	-0.3731071
12	-0.2557397	0.3147796	-0.5481594	-0.7314593
13	0.2535614	-0.9585454	-0.1222306	-0.0442402
14	0.9109831	-0.3336811	0.0331501	-0.2401408
15	-0.7496928	0.3225469	-0.5665945	-0.1135588
16	-0.4710577	-0.3052785	0.4031180	-0.7227765

Table 3: Scores of 10 "respondents" on 16 variables, including measurement and other errors (unique component)

Respondent No.								
1	-0.5182450	1.7459316	0.9071364	0.3686127	2.5799685	0.1209134	0.4857935	-0.8862238
	-1.6103525	0.6112172	-1.0122147	0.5511117	0.5052783	2.2101545	-1.3871889	0.1688346
2	-1.4105225	-0.9148121	-0.1295530	-1.3460293	1.2570162	0.1122990	0.7706858	-0.7391297
	-1.4523983	0.9476829	0.5251331	-1.1696291	0.7196613	0.8456508	-0.6414921	-0.0604200
3	0.0980893	1.5829811	-0.3008118	-0.2126385	1.2071381	1.2388639	-0.1287838	-0.4566827
	0.3116163	-1.0746746	-0.7418182	0.2127893	-1.6790037	-0.3437840	1.0188065	1.2690248
4	0.8240451	0.9567026	0.8769408	0.6688315	0.3058617	0.2077454	0.8710961	-0.7246733
	-0.2282687	-0.4287885	0.4679896	1.3812962	-0.6203752	-1.3698492	-0.0846304	0.5200298
5	0.3234880	2.5084753	1.1316900	-1.7355490	1.1016035	-1.5962896	-0.5106068	1.1926613
	0.1905652	-0.1724078	0.8036264	-0.7793046	-1.2773714	0.2551693	1.7738075	-2.2019291
6	-1.4514360	-0.3160561	-0.3913232	-0.3285195	0.1122166	-0.8328206	-0.5107927	-0.7753966
	-0.2689677	-0.2082363	-0.5259240	0.7332492	0.7330234	0.3067490	0.9486553	0.0424597
7	1.5368481	1.0353651	2.0013475	-0.6179686	0.6170583	0.3557510	-2.0119619	0.6664123
	1.3521461	0.1473606	0.7551325	-0.3877059	-0.3989628	-0.1687852	0.5129723	-0.4095405
8	0.9433647	1.6449528	0.0742216	-1.6763697	0.7742457	-1.5986481	0.6966645	1.4308176
	0.6583975	2.5307007	-0.0285206	1.2261543	-0.7758257	1.1457891	0.1439271	-0.7341754
9	-1.8332605	0.7690830	-1.6466307	-0.2892299	1.0990648	-0.7990375	0.0827256	0.5457472
	0.3168627	-0.7844950	0.3348299	-1.1255226	-1.1724205	-0.2179662	-0.9515268	-0.9126310
10	-0.8479825	0.3257084	0.5361766	1.2312593	1.1899080	-0.9909492	-1.2460098	0.6387758
	-0.3987045	-0.2893403	0.0373909	1.0113802	0.0420041	0.5830306	-0.5677575	-1.1257019

Table 4: Co-efficients to estimate the variables from the factors (factor pattern matrix) including unique component

Var.	Common Factors				Uniqueness Co-efficient
	1	2	3	4	
1	0.4687439	-0.4571099	-0.3863355	-0.3186761	0.5661448
2	0.5006732	-0.4043355	0.1938723	0.2871106	0.6825100
3	0.7199904	-0.3074161	0.1347142	-0.1998271	0.5736116
4	-0.1585296	-0.2694603	-0.6206022	-0.0827186	0.7143319
5	0.4271542	-0.4951494	-0.2936524	-0.0208455	0.6969218
6	-0.5203688	-0.1817660	-0.2522784	-0.1569160	0.7796858
7	0.2181250	0.1260058	-0.6466407	-0.0031189	0.7199931
8	0.7109838	0.2740009	-0.0308322	-0.0241838	0.6464440
9	-0.3034697	-0.5951245	0.1895387	-0.0556619	0.7174329
10	0.3332142	0.3634381	0.3930904	-0.4146335	0.6560792
11	0.3246162	0.1960473	0.0834886	-0.2808726	0.8776845
12	-0.0187992	-0.1538085	0.4033373	-0.4466411	0.7834666
13	-0.4100823	0.5655873	-0.0326667	-0.1635737	0.6957873
14	0.2241583	0.5680439	-0.0745494	-0.2224674	0.7563266
15	-0.2589619	-0.4992898	0.6263841	-0.0916038	0.5318834
16	-0.4241029	-0.2287723	-0.2789811	-0.5768183	0.5977041

Table 5: Correlation Matrix for 16 Variables, based on 200 observations from a 4 factor orthogonal structure with no uniqueness

```
 1.0000000 0.4051787 0.6900114 0.6340237 0.9431350-0.0059713 0.4623582 0.2900544
 0.1004654-0.0996079 0.2080887 0.0174301-0.6948871-0.0726045-0.1786075 0.2275242

 0.4051787 1.0000000 0.8217177-0.2323716 0.6170023-0.6049132-0.2825743 0.4354982
 0.2621879 0.0776430 0.0186103 0.0894879-0.8609477-0.3279745 0.2973240-0.6349148

 0.6900114 0.8217177 1.0000000-0.1082672 0.7172195-0.5364121-0.1169489 0.5982151
 0.0807216 0.4108901 0.3927764 0.3617881-0.8245521-0.0310149 0.1572667-0.3213071

 0.6340237-0.2323716-0.1082672 1.0000000 0.5782025 0.5991549 0.6815043-0.3278183
 0.1945154-0.6829589-0.2797188-0.3652824-0.1731555-0.2272331-0.3003673 0.6312149

 0.9431350 0.6170023 0.7172195 0.5782025 1.0000000-0.0769832 0.3429024 0.2361061
 0.2601505-0.2688802 0.0046583-0.0798047-0.8470567-0.2907568-0.0484546 0.0365543

-0.0059713-0.6049132-0.5364121 0.5991549-0.0769832 1.0000000 0.0082014-0.8996356
 0.5590251-0.5319031-0.6468738 0.1623983 0.1741914-0.4892474 0.2988358 0.9047790

 0.4623582-0.2825743-0.1169489 0.6815043 0.3429024 0.0082014 1.0000000 0.3045617
-0.5691082-0.3067076 0.3501682-0.7149460 0.1965327 0.4860162-0.9007165 0.1541891

 0.2900544 0.4354982 0.5982151-0.3278183 0.2361061-0.8996356 0.3045617 1.0000000
-0.7283376 0.5971771 0.8758367-0.1628014-0.1348697 0.7051781-0.5455699-0.6462631

 0.1004654 0.2621879 0.0807216 0.1945154 0.2601505 0.5590251-0.5691082-0.7283376
 1.0000000-0.4462651-0.8364305 0.4283253-0.5490568-0.9700485 0.8608686 0.3468934

-0.0996079 0.0776430 0.4108901-0.6829589-0.2688802-0.5319031-0.3067076 0.5971771
-0.4462651 1.0000000 0.7701940 0.5829149 0.1168852 0.6001121 0.0106149-0.2383939

 0.2080887 0.0186103 0.3927764-0.2797188 0.0046583-0.6468738 0.3501682 0.8758367
-0.8364305 0.7701940 1.0000000 0.0298213 0.1772743 0.9034441-0.6000730-0.2865623

 0.0174301 0.0894879 0.3617881-0.3652824-0.0798047 0.1623983-0.7149460-0.1628014
 0.4283253 0.5829149 0.0298213 1.0000000-0.2466013-0.2136489 0.7222923 0.3106148

-0.6948871-0.8609477-0.8245521-0.1731555-0.8470567 0.1741914 0.1965327-0.1348697
-0.5490568 0.1168852 0.1772743-0.2466013 1.0000000 0.5576588-0.4206145 0.1584411

-0.0726045-0.3279745-0.0310149-0.2272331-0.2907568-0.4892474 0.4860162 0.7051781
-0.9700485 0.6001121 0.9034441-0.2136489 0.5576588 1.0000000-0.7738513-0.2091888

-0.1786075 0.2973240 0.1572667-0.3003673-0.0484546 0.2988358-0.9007165-0.5455699
 0.8608686 0.0106149-0.6000730 0.7222923-0.4206145-0.7738513 1.0000000 0.1369294

 0.2275242-0.6349148-0.3213071 0.6312149 0.0365543 0.9047790 0.1541891-0.6462631
 0.3468934-0.2383939-0.2865623 0.3106148 0.1584411-0.2091888 0.1369294 1.0000000
```

Table 6: Comparison between Actual and Estimated Factors for structure with 4 Factors and no unique component. Estimates based on SPSS Program

Variable	Factor 1		Factor 2		Factor 3		Factor 4	
	Actual	Estimated	Actual	Estimated	Actual	Estimated	Actual	Estimated
1	0.083	0.049	0.714	-0.603	0.621	-0.717	-0.313	-0.347
2	0.018	0.128	0.925	-0.929	-0.144	0.060	0.350	0.341
3	0.234	0.299	0.952	-0.927	-0.017	-0.056	-0.199	-0.220
4	-0.287	-0.373	0.319	0.086	0.947	-0.912	-0.139	-0.146
5	-0.052	-0.056	0.822	-0.737	0.566	-0.670	-0.368	-0.060
6	-0.761	-0.824	-0.430	0.403	0.322	-0.230	-0.365	-0.325
7	0.466	0.380	-0.178	0.301	0.866	-0.875	0.031	0.004
8	0.908	0.938	0.417	-0.341	-0.015	-0.057	0.029	0.008
9	-0.930	-0.912	0.352	-0.393	-0.101	0.119	0.004	0.018
10	0.604	0.619	0.144	-0.154	-0.564	0.565	-0.544	-0.523
11	0.922	0.923	0.104	-0.034	-0.017	-0.013	-0.373	-0.383
12	-0.256	-0.258	0.315	-0.346	-0.548	0.583	-0.731	-0.687
13	0.254	0.215	-0.959	0.956	-0.122	0.198	-0.044	-0.036
14	0.911	0.890	-0.334	0.382	0.033	-0.039	-0.240	-0.248
15	-0.750	-0.693	0.323	-0.409	-0.567	0.588	-0.114	-0.087
16	-0.471	-0.581	-0.305	0.337	0.403	-0.316	-0.723	-0.671
Correlation between Actual and Estimated	0.995		-0.993		-0.991		0.997	

Table 7: Alternative Factor Pattern to generate the correlations shown in Table 5

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
VAR001	0.14714	0.93331	-0.18453	0.27075
VAR002	-0.17856	0.69088	0.15299	-0.68361
VAR003	0.22210	0.86130	0.33065	-0.31558
VAR004	-0.19616	0.39915	-0.60678	0.65866
VAR005	-0.09337	0.96514	-0.24075	0.04332
VAR006	-0.51532	-0.17328	-0.02401	0.83891
VAR007	0.43841	0.17723	-0.83148	0.29184
VAR008	0.78173	0.32661	-0.06936	-0.52665
VAR009	-0.88479	0.26011	0.33882	0.18628
VAR010	0.71469	-0.03299	0.67485	-0.18251
VAR011	0.97947	0.12390	0.07200	-0.14174
VAR012	-0.02445	0.14388	0.95999	0.23553
VAR013	0.33962	-0.89715	-0.16918	0.22609
VAR014	0.95693	-0.24373	-0.14945	-0.04998
VAR015	-0.67865	0.06763	0.73103	-0.02572
VAR016	-0.15543	-0.00191	0.07800	0.98513

Table 8: Average CPU Time taken by three programs to analyse a 16 variable, 200 observation data set

Program	Average CPU Time (minutes)	Number of Runs on which average is based
SPSS	2.06	3
PCVARIM	1.52	3
SSP*	3.75	2

* Estimate, based on time taken to extract and rotate four factors

Table 9: Average Correlations between actual coefficients and those estimated by three Programs for samples of 200 observations from 2 factor structures. (Both factor structures are orthogonal and contain 4 factors. One contains a unique component but the other does not.)

Program	Average correlation between actual and estimated coefficients for structure without uniqueness	Average Correlation between actual and estimated coefficients for structure with uniqueness
SPSS	0.99	0.89
PCVARIM	0.99	0.88
SSP	0.99	0.88

Table 10: Five-, four- and three-factor solutions estimated from 200 observations on structure given in Table 4

FIVE- FACTOR SOLUTION

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
VAR001	0.71102	-0.26056	0.03556	-0.31505	0.05437
VAR002	0.38439	-0.42123	-0.44927	0.37219	-0.03579
VAR003	0.61796	-0.16022	-0.53588	-0.25263	-0.05006
VAR004	0.35674	-0.27825	0.46740	-0.16515	-0.15668
VAR005	0.69855	-0.30252	0.13737	0.02938	-0.11198
VAR006	-0.13775	-0.22166	0.48328	-0.25432	-0.00545
VAR007	0.51278	0.23624	0.48642	-0.09791	0.05083
VAR008	0.60179	0.48637	-0.22978	0.05983	0.33683
VAR009	-0.19709	-0.65465	-0.12061	-0.22462	0.28155
VAR010	-0.04914	0.48657	-0.52904	-0.35252	-0.06792
VAR011	0.25813	0.40503	-0.20925	-0.28836	-0.13443
VAR012	-0.16679	-0.00117	-0.35144	-0.40745	-0.00370
VAR013	-0.52727	0.48886	0.16769	-0.15290	0.07255
VAR014	0.10410	0.66586	0.05450	-0.15628	-0.08873
VAR015	-0.50701	-0.51037	-0.44393	-0.13958	-0.10109
VAR016	-0.10641	-0.30275	0.23877	-0.60503	0.06479

FOUR- FACTOR SOLUTION

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
VAR001	0.72798	-0.21368	0.03483	-0.31511
VAR002	0.40700	-0.40552	-0.44686	0.36965
VAR003	0.62863	-0.12522	-0.54106	-0.25334
VAR004	0.37151	-0.24672	0.45982	-0.16236
VAR005	0.71447	-0.25544	0.13539	0.02874
VAR006	-0.12332	-0.22563	0.48987	-0.25881
VAR007	0.50255	0.27663	0.48159	-0.09387
VAR008	0.53931	0.48995	-0.21839	0.05956
VAR009	-0.15452	-0.64200	-0.10284	-0.21231
VAR010	-0.07866	0.48229	-0.53648	-0.34746
VAR011	0.23340	0.41977	-0.21527	-0.27919
VAR012	-0.16808	-0.01348	-0.35170	-0.40950
VAR013	-0.55528	0.45863	0.16480	-0.14943
VAR014	0.06663	0.67739	0.04426	-0.14876
VAR015	-0.47964	-0.54949	-0.43432	-0.14705
VAR016	-0.08764	-0.30647	0.24488	-0.61325

THREE- FACTOR SOLUTION

	FACTOR 1	FACTOR 2	FACTOR 3
VAR001	0.69389	-0.23058	0.01454
VAR002	0.36024	-0.40589	-0.40548
VAR003	0.60418	-0.15690	-0.54805
VAR004	0.37071	-0.25253	0.45990
VAR005	0.71726	-0.28905	0.13061
VAR006	-0.11957	-0.20464	0.48398
VAR007	0.52592	0.26858	0.46956
VAR008	0.55985	0.47190	-0.25076
VAR009	-0.17633	-0.62766	-0.08686
VAR010	-0.06579	0.45350	-0.52074
VAR011	0.24088	0.39864	-0.23202
VAR012	-0.16356	-0.00928	-0.33113
VAR013	-0.53627	0.48923	0.16613
VAR014	0.09256	0.67892	0.02367
VAR015	-0.50599	-0.53877	-0.41459
VAR016	-0.08091	-0.25388	0.20376

Table 11: Original Co-efficients and those estimated from (a) a dichotomized data set with equal endorsement and (b) a dichotomized data set with unequal endorsement

VARIABLE NO.	ORIGINAL CO-EFFICIENTS
1	0.0832195 0.7136061 0.6211666-0.3130382
2	0.0183298 0.9254015-0.1436086 0.3502466
3	0.2339566 0.9515763-0.0166898-0.1987171
4	-0.2874494 0.0319199 0.9471081-0.1390697
5	-0.0519799 0.8221784 0.5656589-0.0367599
6	-0.7611248-0.4298670 0.3215278-0.3640375
7	0.4660683-0.1779892 0.8661069 0.0309699
8	0.9083121 0.4170263-0.0153999 0.0286597
9	-0.9303820 0.3524770-0.1006791 0.0037600
10	0.6035873 0.1443493-0.5643376-0.5443977
11	0.9217334 0.1044291-0.0171199-0.3731073
12	-0.2557397 0.3147796-0.5481594-0.7314593
13	0.2535614-0.9585454-0.1222306-0.0442402
14	0.9109833-0.3336811 0.0331501-0.2401408
15	-0.7496928 0.3225469-0.5665945-0.1135588
16	-0.4710577-0.3052785 0.4031180-0.7227765

Co-efficients estimated from a data set with equal endorsement

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
VAR001	-0.51439	0.46013	-0.48947	0.20361
VAR002	-0.50031	0.59148	0.01160	-0.27387
VAR003	-0.37602	0.75610	-0.06487	0.08087
VAR004	-0.39207	-0.29329	-0.59966	0.18667
VAR005	-0.61309	0.45499	-0.47052	0.06475
VAR006	-0.32736	-0.72372	-0.06529	0.38679
VAR007	0.18036	-0.17608	-0.80284	-0.00579
VAR008	0.33439	0.74681	-0.12820	-0.10653
VAR009	-0.81343	-0.21643	0.21235	-0.02372
VAR010	0.30878	0.52517	0.31161	0.45430
VAR011	0.54943	0.56439	-0.13776	0.36594
VAR012	-0.17990	0.31014	0.50976	0.52229
VAR013	0.69309	-0.54671	0.04803	0.04815
VAR014	0.79610	0.27326	-0.16075	0.23731
VAR015	-0.51988	0.05628	0.62807	0.03188
VAR016	-0.30650	-0.40476	-0.12239	0.64640
Highest corr with orig. Factor No.	1	2	3	4
Correlation	0.80	0.76	-0.94	-0.97

Co-efficients from data set with unequal endorsement

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4
VAR001	0.77750	-0.11180	-0.24306	-0.19935
VAR002	0.74417	0.04122	0.21408	0.19710
VAR003	0.79723	0.26241	0.12093	-0.01079
VAR004	0.20849	-0.56464	-0.54069	-0.09482
VAR005	0.60338	-0.25166	-0.22659	-0.02107
VAR006	-0.26020	-0.60345	-0.07506	-0.35344
VAR007	-0.00150	-0.03671	-0.65776	0.01792
VAR008	0.29152	0.51498	-0.31207	-0.00076
VAR009	0.12963	-0.39936	0.14077	-0.38268
VAR010	-0.08301	0.59995	0.10087	-0.45175
VAR011	0.06053	0.55615	-0.36564	-0.28872
VAR012	0.14603	0.21705	0.35877	-0.49245
VAR013	-0.39140	0.15569	-0.19631	-0.04845
VAR014	-0.26050	0.36522	-0.39529	-0.20257
VAR015	0.13207	-0.18503	0.24528	-0.33730
VAR016	-0.08876	-0.29441	-0.00922	-0.36044
Highest corr with orig. Factor No.	2	1	3	4
Correlation	0.94	0.84	-0.75	0.80