ERLING BERGE

ON THE CORRELATION OF FACTORS IN FACTORIAL ECOLOGY

NOTAT
87:12 *Institutt for sosialforskning, Munthes gt. 31, 0260 Oslo 2*

ON THE CORRELATION OF FACTORS IN FACTORIAL ECOLOGY

Erling Berge
Institute of Applied Social Research
Munthesgate 31
OSLO 2, Norway

ABSTRACT

Most studies in factorial ecology use orthogonal factors. But the assumption is that this, at best can be considered a theoretically limiting case. In "reality" factors are assumed to correlate.

Introducing a distinction between factors describing the structure of a social system and factors describing the environment of the social system, the present paper argues that factors describing the structure of a social system in most cases will be found to be orthogonal. They will, however, have to correlate with factors describing the environment of the social system. A reanalysis of a previous study of the Norwegian factorial ecology gives some support for the argument.

ON THE CORRELATION OF FACTORS IN FACTORIAL ECOLOGY*

Factorial ecology typically starts out with a universe of spatially defined units for which there are defined a set of variables. The variables are defined with the aim of gaining a comprehensive description of the resources and living conditions of the population within each unit. Routinely this includes a description of land and population size, demographic characteristics, industrial composition, occupational mix and educational statuses of the population, as well as its housing conditions, income distribution and political preferences.

The analysis of such variables in factor models usually assumes uncorrelated dimensions. The initial argument for assuming uncorrelated factors seems mostly to have been technical: the mathematics is much simpler and the computational procedures possible to do by hand. There also was - and still is - a certain appeal in the parsimony and mathematical elegance it provides. But mathematical elegance must not blind us to the real world: "All experience of rotation alike with data on physical, biological, or social science, forces upon us the truth that in nature factors are correlated." (Cattell, 1952,pp.117.) More or less this statement seems to cover the theoretically reflected judgments of social scientists today (see f.i. Coleman 1964, Hunter 1972, and Hamm 1979). Uncorrelated factors are at most to be considered as a theoretically limiting case.

This is a slightly revised version of a paper originally presented to the Xth World Congress of Sociology, Research Committee on Social Ecology, Mexico City, Mexico, August 1982. I appreciate the comments received there. In particular I wish to thank Frank L. Sweetser, who also suggested the topic of the paper.

However, recent studies (Hamm 1979, Berge 1981) show a remarkable robustness of the main factor dimensions across both different methods of factorization and different degrees of correlation allowed between factors extracted. It would seem that the orthogonal solutions usually employed, in most cases not only give a theoretically meaningful description of the social ecological differentiation of the analytical units, but in certain respects also give a better description than oblique factors.

Abu-Lughod (1969) has tried to outline the conditions which are likely to produce uncorrelated factors. Both specialization of actors and of land use contribute to a development where it will be increasingly likely to find independence among factors in social ecological studies. Independent dimensions is a sufficient condition for finding orthogonal factors, but it is not a necessary condition. Uncorrelated factors can not be interpreted as independent factors (Janson 1969, Johnston 1971). It has, for instance, been pointed out that life cycle factors which by their very nature have to be curvelinearly interrelated (Janson 1969, 1980), very well may be represented by uncorrelated factors.

It may be that it is the correlated factors which are in need of a theoretical defence. Why do one sometimes have to employ oblique factors in order to arrive at a meaningful description of a social ecological system?

The conclusion of Sweetser (1974) to combine orthogonal and oblique factors may be the practical advice to follow. But is there any way to predict which factors are to be oblique while others are orthogonal?

The discussion of oblique vs. orthogonal factors in the litterature does not offer much help. But Janson (1980, pp. 446) concludes that "On the community level oblique systems are preferable if both urbanism and size are to be

given a chance to come forward at full strength." This may be a clue.

Theoretically considered there is a basic difference between "urbanism" and "size". While urbanism may be interpreted to say something about the social structure of the society, size may be saying something about the scale of the society, or perhaps better; the environment of the social system.

We shall see that a distinction between social system and environment shall prove fruitful for the present problem. The present paper will go into the problem of correlation among factors in factorial ecology by proposing a simple model of a social ecological system. The model will explain which kind of factors one ought to expect to correlate with a "size" factor, or more generally with environmental factors.

A social eco-system.

A simple model of a social eco-system might distinguish between the social system proper and the environment of the system (f.i. the habitat of the population).

Factorial ecology as described above takes this environment, divides it into suitable spatial units and proceeds to characterize these and the populations they contain. A distinction between variables describing the environment and variables describing the social system is not utilized.

Yet, if one regards the problem of interdependence between a social system and its environment it seems fairly obvious that the environment must represent constraints which influence the structure of the social system.

If one conceptualizes the social system as consisting of a social structure which social processes are working to reproduce or transform, the environment must influence the shape of both. The members of a social system adapt to its habitat and its particular distribution of natural resources by shaping the social processes of the system to take advantage of the existing conditions and counteract the continous flow of effects from the natural processes (seasons, weather, disasters, diseases).

In factor analytic studies some variables describe the environment and some describe the social system. It seems reasonable to expect that some factors ought to describe the environment and some the social system. Direct data on the social processes are usually missing. Indirect data like change indicators are seldom used. Therefore the data describing the social system usually refer to aspects of the social structure.

The factors defined by such variables must accordingly be interpreted as a description of the social structure of the system.

The specialization of actors and the differentiation of activities according to location make it likely that the basic factors describing a social structure will appaer as uncorrelated factors. But these factors can not be expected to be uncorrelated with the factors describing the environment of the structure.

While our knowledge of social structure and its spatial distribution lead us to expect uncorrelated factors describing the structure, we do not know much about which factors to expect in a study of the environment or how they may interrelate.

The variables describing the environment of the social system may either be direct measures of the distribution of natural resources and geographical features of the units of analysis or indirect measures of these based on their impact on the human activittes within the units. Considered by themselves the environmental factors do not seem to be more than weakly interrelated (climate f.i. will be somewhat related to geographical features). But the way boundaries are drawn around the units of analysis will confound this picture. In particular this happens if our measurement of the factors have to rely on indirect indicators like population density or land area which are so closely related to the way boundaries are drawn and which often also are taken into consideration when boundaries are defined. This must be accounted for in a study of environmental factors.

The sentral proposition in this paper is, however, the existence of environmental factors and that environmental factors and social factors have to intercorrelate in a meaningful way.

A reanalysis of data from a traditionally designed factor analytic study of Norwegian Communes will be used to test these propositions.

Results.

The data used have been described in Berge (1981), and only a short outline will be given here.

Data on the 451 Norwegian communes as of 1. January 1970 were collected from the Population and Housing Census of 1970 *) and other sources. Neighboring communes were agg-

*)I am grateful to the Central Bureau of Statistics of Norway, and to the Norwegian Social Science Data Services for making data available for the study.

regated to reach a minimum population size of 500. This resulted in 448 analytical units. For each unit a total of 113 analytical variables were defined and computed (per cent variables, ratios, indexes). To reduce skewness and kurtosis logarithm and square root transformations were used. Of the 113 variables 41 are used in the present study. Their definitions and transformations used are listed in Appendix tables A1 and A2.

Of the inital 113 variables 11 may be said to be mainly determined by environmental characteristics. For these 11 variables a separate factor analysis was undertaken resulting in two environmental factors defined by 7 variables. Two variables had to be excluded because of too high intercorrelations with other variables.*)

The variables excluded were no. 7 "Mean size of agglomerations" because of a correlation of .993 with variable no. 2 "Number of people in densely settled areas", and no. 5 "Inhabitants per km^2" because of a correlation of -.835 with variable no. 1 "Land area".

Two more variables (no.9 and 10 in Table A1) had to be excluded since they did not have any intercorrelations with other variables in the matrix as high as .5 (see Sweetser 1974 for practical guidance to factor analysis of ecological variables).

---

*)*Exclusion of variables with high intercorrelations is necessary if factor scores are to be computed. High intercorrelations mean a high degree of linear dependency in the matrix. The determinant of the correlation matrix will be close to zero, and the computed factor scores will be inaccurate because of rounding errors or impossible because of zero division. However, knowledge of the intercorrelations will certainly help the interpretation of the factors arrived at (see Berge 1981).*

The analysis of the remaining seven variables resulted in two factors. In order to test the possibility of intercorrelations between them, four rotations were done, one orthogonal according to the varimax criterion, and three oblique according to the oblimin criterion with DELTA set to .5, .0, and -.5 (see Table A4 and A5). The definitions of the factors seem very much the same in all rotations. And the correlation coefficients between factors from the orthogonal solution and the oblique solution with DELTA= .0 are as high as .98.

The environmental variables as measured by the available data seem to be adequately described by the two orthogonal factors.

The two factors are interpreted to represent a LAND SIZE factor and a POPULATION SIZE factor. The factor matrix, slightly rearranged is as follows (see also Table A4 and A5):

VARIABLE		FACTOR LOADINGS	
NO	NAME	POPULATION SIZE	LAND SIZE
2	No. of people in densely settled areas	.91	-.14
4	Total number of people	.77	-.24
8	% of the population in densely settled areas	.75	-.38
6	No. of agglomerations	.68	-.10
1	Total land area in km^2	.01	.65
11	Dairy farms in % of all farms	-.43	.77
3	Farms with 10 ha in % of all farms	-.43	.82

The labels of the factors need some qualifications. The LAND SIZE factor obviously is tied in with the conditions for agriculture. Perhaps "arable land" might be a better label. The factor thus tells something of how the environment is suited for agricultural activities. Likewise it may be seen that the POPULATION SIZE factor is tied in with population density. This factor may then tell something about the conditions for certain kinds of human activites. Most particulary those associated with urban societies.

Of the 113 variables defined in Berge (1981) 60 were found suitable for inclusion into factor analysis. These 60 variables defined 6 factors labeled SOCIO-ECONOMIC STATUS; FAMILISM, DEPRIVATION, AFFLUENCE, MANUFACTURING INDUSTRY, and FEMALE ECONOMIC ACTIVITY. By successive removal of variables it was found that 30 variables were sufficient to define the six factors. The coefficients of correlation between factors from the 60 variable solution and the 30 variable solution varied from .95 to .98 (correlation of factor scores). The variables defined in table A2 are the same as those in the original 30 variable solution except for two changes. Since the variables "% farms with 10+ da." and "Inhabitants pr km^2" were among the variables taken to describe the environment, they were replaced by "Dependent on agriculture" and "Income of 60000+" (variables no 19 and 27 in table A2). In table A3 the factor matrix of the analysis of the 30 variables is reported. Correlation of factor scores for the six factors used here and the six original factors gives coefficients ranging from .97 to 1.00.

The main question addressed here, however, is whether the factors describing the environment of the social system will correlate with the factors describing the structure of the social system.

Coefficient of correlation between factors describing system environment and system structure.

	Population Size	Land Size
Socio-economic status	.46	-.44
Familism	.23	.06
Deprivation	.35	.10
Affluence	.27	-.43
Manufacturing industry	.31	-.32
Female economic activity	.00	.28

The coefficients above are not very high. Only three higher than .4, and two more are between .3 and .4. But the pattern seems to be what one might have expected.

Recalling that LAND SIZE mostly means arable land size and that POPULATION SIZE also has aspects of density, it is not surprising that SOCIO-ECONOMIC STATUS is the one structural factor most affected by the environmental factors and FAMILISM the one least affected. Likewise it is known that both affluence and relative deprivation are most clearly present in the larger cities and that manufacturing industry means some kind of agglomeration. It is, however, worth noting the low correlation of POPULATION SIZE and MANUFACTURING INDUSTRY. This would seem to be in accord with the observation that much manufacturing industry has moved out of the larger agglomerations. The relation between FEMALE ECONOMIC ACTIVITY and LAND SIZE is an insteresting affirmation of the relatively large impact from the inclusion of female family labor on farms into the stock of economically active women.

The most interesting observation here may, however, be the relation between SOCIO-ECONOMIC STATUS and POPULATION SIZE. Among the main characteristics of the urbanization process is the growth of population and the increasing density. But urbanization has come to mean much more than

that. In Norway for example the close correlation of variables indicating SES and variables indicating urbanization has led to conceptual confusion of the two. They have sometimes been used interchangably. The separation of variables into those describing the system environment and those describing the social system separates the two concepts and takes care of the interrelation by allowing a SES factor and a URBANIZATION factor to correlate.

Urbanization here means only size and density of population. This may be thought of as an environmental characteristic of a social system in the sense that size and density is something the actors have to take into consideration in all their actions: it shapes their choice of activities and thus shapes the social structure. But obviously size and density of a population also is a result of the impact social activities has on the environment. As material infrastructure (building, roads, etc.) acumulate, the environment changes.

Using a rather different approach Sweetser (1982) arrives at a very similar conclusion in a study of Urban Residental areas in Australia. Comparing Urban and Rural residential areas he finds that "there appear to be two district modes of directional differentiation, one associated with changes in urban community size, and the other with the shift from urban to rural communities" (pp.154). The distinction between city size on the one hand and a rural-urban shift on the other would seem to be a close approximation to what I have called the environmental factors of population size and land size.

The boundary between a system and its environment can not be a fixed line. Like so much else it has to be defined in relation to the problem investigated. If population size (density) and land size (arable) are considered as part of the environment of the social system and not as belonging

to the social system, the reanalysis of our data suggests that environmental factors exist and that they correlate as one might have expected with factors describing the structure of the social system.

REFERENCES

ABU-LUGHOD, J.L. (1969) "Testing the theory of social area analysis: The ecology of Cairo, Egypt.", AMERICAN SOCIOLOGICAL REVIEW, 34:198-212.

BERGE, E. (1981) "The social ecology of human fertility in Norway 1970. Ann Arboar, University Microfilms.

CATTELL, R.B. (1952) "Factor analysis." New York, Harper Brothers

Coleman, J.S. (1964) "Introduction to Mathematical Sociology", Glencoe, Free Press.

HAMM, B. (1979) "Indikatoren der Stadtentwickelung". Trier,Universitat Trier.

HUNTER, A.A. (1972) "Factorial ecology: a critique and some suggestions". DEMOGRAPHY 9:107-117.

JANSON, C. G. (1969) "Some problems of ecological factor analysis", in QUANTITATIVE ECOLOGYCAL ANALYSIS IN THE SOCIAL SCIENCES" eds. M. Dogan and S. Rokkan, pp.301-41. Cambridge, Mass.,MIT Press.

JANSON, C. G. (1980) "Factorial Social Ecology: An Attempt at Summary and Evaluation", in ANNUAL REVIEW OF SOCIOLOGY 6:433-56, Palo Alto, Annual Reviews, Inc.

JOHNSTON, R.J. (1971) "Some limitations of factorial ecology and social area analysis". pp. 314-23, ECONOMIC GEOGRAPHY, June(Suppl.)

SWEETSER, F.L. (1974) "The uses of factorial ecology in classification", in CURRENT RESEARCH IN SOCIOLOGY, ed. M. S. Archer, pp. 317-43, The Hague, Mouton.

SWEETSER, F.L. (1982) Urban Residential Areas in Australia. Departmental Monograph No. 4, Canberra, Department of Sociology, The Australian National University.

Table 1A. ECOLOGICAL VARIABLES FOR THE STUDY OF SPATIAL DIFFERENTIATION OF SOSCIAL STRUCTURE IN NORWAY 1970: 11 SYSTEM ECOLOGICAL VARIABLES:

VARIABLE NO		DEFINITION	TRANSFORMATIONS
1	(1)*	The total land area of the commune in km^2	Log.transformation
2	(2)	The absolute number of people living in densely settled areas	Log.transformation
3	(3)	The % of all farms having more than 10 dekar arable land	
4	(4)	The total number of people	Log.transformation
5	(6)	The nummber of inhabitants per km^2 land	Log.transformation
6	(7)	The number of agglomerations in the commune	
7	(8)	The mean population size of the agglomerations	Log.transformation
8	(9)	The % of the population living in densely settled areas	
9	(10)	The % of all farms with more than 10 dekar which have 20-75 dekar arable land	
10	(11)	The % of all forest properties which are less than 250 dekar in size	
11	(82)	The % of all farms which are dairy farms	

*)No. from Table 1, Appendix A in Berge 1981

Table 2A. ECOLOGICAL VARIABLES FOR THE STUDY OF SPATIAO DIFFERENTITATION OF SOCIAL STRUCTURE IN NORWAY 1970: 30 SOCIAL ECOLOGICAL VARIABLES.

VARIABLE NO		DEFINITION	TRANSFORMATION
1	(12)*)	The % of the population of age 5 to 14 years	
2	(13)	The % of the population of age 65 years or more	
3	(14)	The % of the population aged 20 to 59 years who are 20 to 39 years	
4	(32)	The % of the families with more than 1 person who have 4 or more unmarried children	
5	(34)	The % of all households which have unmarried children and both parents	
6	(39)	The % of all occupied housing units which have more than 1.0 persons per room	
7	(42)	The % of all children of age 0 to 14 who live in private housing units with more than 1.0 persons per room	
8	(43)	The % of all men older than 15 years who have their own housing unit	
9	(51)	The % of all housing units which are in one family structures	
10	(52)	The % of all housing units which are in farm houses	
11	(54)	The % of all housholds which have at least 5 rooms	
12	(55)	The % of all households which have thelephone	
13	(59)	The % of all persons of age 16 or more who are occupied within commune of residence	
14	(62)	The % of the women aged 16 to 59 who have children in the age group 0 to 12 years and who are economical active	
15	(63)	The % of the women aged 20 til 59 who are economically active	
16	(68)	The % of all men aged 16 or more who are occupied in professional or managerial occupations (occupational codes 00-33, 80-89)	
17	(69)	The % of all men aged 16 or more who are occupied in blu-collar occupations (occupation codes 50-59, 70-89)	

*)*No. from Table 1. Appendix A in Berge 1981*

Table 2A continued: 30 SOCIAL ECOLOGICAL VARIABLES, NORWAY 1970

VARIABLES NO		DEFINITION	TRANSFORMATION
18	(76)*)	The number of persons aged 16 or more with main income from work in services (industry codes 811-93) per 100 persons with main income from manufacturing (industry codes 2-3)	Square root
19	(77)	The % of the total population who are dependent on agriculture for their main income (industry codes 01-02)	
20	(78)	The % of the total population who are dependent on manufacturing for their main income (industry codes 11-39, 51-52).	
21	(79)	The % of the total population who are dependent on trade for their main income (industry codes 61-66)	
22	(81)	The number of pensioners per 100 persons economically active	
23	(86)	The mean number of workers employed per corporation in manufacturing (industry codes 20-39)	Square root
24	(94)	The % of all voters casting their vote for the Labor Party (AP), Socialist Peoples Party (SF) and the Communist Party (K). Storting election 1969	
25	(98)	Tax to the commune in kr. in 1968 per inhabitant in the commune as of 1.1.1968	Square root
26	(99)	Transfers from the state to cover expenditures in the cultural, educational and welfare sectors in kr. per inhabitant aged 16 or more at the end of 1970	
27	(100)	The % of all personal tax payers with taxable income of kr. 60.000,- or more	Square root
28	(106)	The % of all aged 25-69 who have primary education only	
29	(107)	The % of all aged 25-69 who have education at gymnasium level II of III	
30	(111)	The number of cars per 100 families	

*)*No. from Table 1. Appendix A in Berge 1981*

Table 3A. DIMENSIONS OF THE NORWEGIAN SOCIAL STRUCTURE IN 1970. 30 VARIABLES ON 448 UNITS OF 451 COMMUNES. VARIMAX ROTTED FACTOR MATRIX OF A PRINCIPAL FACTORS SOLUTION

VARIABLES NO	SHORT NAME	FACTOR COEFFICIENTS SOCIO ECONOMIC STATUS	FAMIL-ISM	DEPRI-VATION	MANUFAC-TURING INDUSTRY	AFFLU-ENCE	FEMALE ECONOMIC ACTIVITY
1	% age 5-14	-.03	.83	-.09	.00	-.37	-.12
2	% age 65+	-.31	-.82	-.37	-.14	.00	.10
3	% age 20-39 of 20-59	.38	.73	.29	.08	-.01	.08
4	Large Families	-.23	.39	-.11	-.08	-.79	.05
5	% child families	.12	.89	-.01	.18	.07	-.21
6	Housing units 1.01+person	.35	.38	.80	.10	-.03	.16
7	Children in HU´s 1.01+person	.21	.13	.81	.11	-.16	-.08
8	% men with own dwelling	.34	.09	.02	.23	.72	-.03
9	% HU´s in one family stru.	-.16	.15	.04	-.01	-.11	-.72
10	% HU´s in farm houses	-.62	-.21	-.45	-.29	-.20	.31
11	% households with 4+ room	-.39	-.04	-.77	-.17	-.34	-.03
12	% HH´s with telephone	.14	-.13	-.68	-.19	-.02	.24
13	% occupied within commune	-.26	-.15	-.09	.06	-.07	.64
14	% ec.act.women with child	-.17	.33	-.16	-.25	-.10	.70
15	% women age 20-59 ec.act.	.03	-.03	-.12	-.16	.19	.87
16	% men in prof./manag.occ.	.89	.11	.10	-.04	.17	-.05
17	% men in blue collar occ.	.04	.04	.29	.72	.37	-.03
18	% Rate occ.in serv./manufac.	.01	-.08	.05	-.65	-.16	.16
19	% dependent on agric.	-.61	-.11	-.40	-.37	.02	.39
20	% dependent on manufac.	.14	.19	.18	.90	.27	-.04
21	% dependent on trade	.68	.25	.24	-.02	.44	.04
22	Rate pensioners/ec.active	-.24	-.73	-.21	-.18	-.32	-.23
23	Mean no.workers per firm	.32	.11	.22	.65	.14	.03
24	% votes for left parties	-.12	-.08	.78	.01	.17	-.27
25	Commune tax per capita	.49	.07	.15	.31	.66	.13
26	State transfers per redid.	-.29	.09	.05	-.35	-.71	-.03
27	% tax payers inc. 60.000+	.71	.19	.03	.25	.33	-.08
28	% with primary school	-.50	-.18	.31	-.13	-.43	-.44
29	High education	.87	.16	-.04	.14	.33	.07
30	No of cars per family	.08	.06	.04	.22	.70	.15
FACTOR VARIANCE		4.9	4.0	4.0	3.1	4.0	3.0

Table 4A. POPULATION SIZE: CORRELATIONS BETWEEN VARIABLES AND FACTOR.
PRINCIPAL FACTORS MODEL: four rotations

VARIABLES NO	short name	ORTHOGONAL VARIMAX	OBLIQUE OBLIMIN DELTA = .5	OBLIMIN DELTA = .0	OBLIMIN DELTA = -.5
1	Totalland area in km^2	.01	-.18	-.14	-.13
2	Total pop. in densely settled areas	.91	.91	.92	.92
3	% of farms with 10+ da	-.43	-.65	-.60	-.60
4	Total population	.77	.81	.81	.81
6	No of agglomerations	.68	.68	.69	.69
8	% of pop. in densely settled areas	.75	.83	.82	.82
11	% dairy farms	-.43	-.64	-.59	-.59
Correlation between density and size		r=.0	r=-.58	r=-.40	r=0.34

Table 5A. LAND SIZE: CORRELATIONS BETWEEN VARIABLES AND FACTOR.
PRINCIPAL FACTORS MODEL: four rotations

VARIABLES NO short name	ORTHOGONAL VARIMAX	OBLIQUE OBLIMIN DELTA= -.5	OBLIMIN DELTA = .0	OBLIMIN DELTA = .5
1 Total land area in km^2	.65	.64	.64	.62
2 Total pop.in densely settled areas	-.14	-.25	-.31	-.41
3 % of farms with 10+ da.	.82	.86	.88	.91
4 Total population	-.24	-.33	-.37	-.46
6 No of agglomerations	-.10	-.19	-.23	-.31
8 % of pop. in densely settled areas	-.38	-.47	-.52	-.60
11 % dairy farms	.77	.82	.84	.87
Correlation between density and size	r=.0	r=-.34	r=-.40	r=-.58