DOCUMENT RESUME

ED 082 065 CE 000 381

AUTHOR Beck, E. M.
TITLE A Canonical Approach to Assessing Occupational Mobility Matrices.
INSTITUTION Wisconsin Univ., Madison. Center of Applied Sociology.
SPONS AGENCY Economic Development Administration (DOC), Washington, D.C. Office of Economic Research.; National Inst. of Mental Health (DHEW), Rockville, Md.
REPORT NO RID-73-6
PUB DATE Jan 73
NOTE 19p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS *Correlation; Economic Change; *Economic Development; *Matrices; Mobility; *Occupational Mobility; Relocation; Socioeconomic Status

ABSTRACT

A canonical technique is explored that permits assessment to which origin positions determine destination positions within occupational structures. This procedure requires expressing the occupational positions as binary variables, then obtaining canonical correlations among the sets of binary variables. The association between origin and destination positions is taken as the mean square of these canonical correlations. An application of the technique to intra-generational mobility during regional economic development is provided. (Author)

CENTER OF APPLIED SOCIOLOGY

University of Wisconsin-Madison

Cooperative Extension Programs

University of Wisconsin-Extension

A CANONICAL APPROACH TO ASSESSING OCCUPATIONAL MOBILITY MATRICES*

January, 1973

Working Paper RID 73.6

E. M. Beck
Institute of Behavioral Science
and Department of Sociology
University of Colorado
Boulder, Colorado 80302

* Preparation of this paper was partially supported by the College of Agriculture and Life Sciences at the University of Wisconsin, Madison; PHS Grant MH-19689 from the National Institute of Mental Health, Gene F. Summers and John P. Clark, Co-Principal Investigators; and the Office of Economic Research, Economic Development Administration, Grant OER-417-G-72-7 (99-7-13248), Gene F. Summers, Principal Investigator.

ABSTRACT

A canonical technique is explored that permits assessment to which origin positions determine destination positions within occupational structures. This procedure requires expressing the occupational positions as binary variables, then obtaining canonical correlations among the sets of binary variables. The association between origin and destination positions is taken as the mean square of these canonical correlations. An application of the technique to intra-generational mobility during regional economic development is provided.

A CANONICAL APPROACH TO ASSESSING OCCUPATIONAL MOBILITY MATRICES

Background

The analysis of occupational mobility is a recurrent topic for sociological investigation. Most studies of mobility have concentrated on assessing the amount of movement between different positions within the occupational structure over a discrete period of time (cf. Glass, 1954; Rogoff, 1953; Lipset and Bendix, 1959; Blumen, et al., 1955). Often this assessment is made by comparing the observed mobility rates between social origins and destinations against those expected if the destination positions were independent of the origins. In other words, the actual rate of mobility is compared to the rate anticipated if the origin positions had no systematic effects on determining the destinations (Bartholomew, 1967). Departures away from independence are then interpreted as the effects of "inheritance". There seems to have been a tendency to equate the effects of inheritance with the lack of mobility between origin and destination positions (cf. Rogoff, 1953). Occupational systems which demonstrate considerable mobility are then held to manifest minimal inheritance, and be relatively open.

The position taken here is that it may be more reasonable, or at least as conceptually viable, to view inheritance as the degree to which an origin position determines the destination position. This may, or may not, imply a lack of mobility. The distinction we wish to make here closely parallels the difference in focus between classical mobility research and investigations of the status attainment process (Carter and Carter, 1971). An illustration adapted from Klatzky and Hodge (1971) will make the point clear. Consider a society in which every "blue collar" worker became, automatically through some unknown process, "white collar" over a discrete interval of time, and every "white collar" worker became "blue collar". Such a social system would be characterized as being highly mobile, yet in this hypothetical case, the mobility is perfectly determined and predictable from knowing the worker's position in the occupational structure prior to the period of transition. It is entirely possible to conceive of an occupational structure in which there is complete mobility, yet there is absolute inheritance in the sense that every move is completely determined by the origin position. While the amount of mobility and the lack of inheritance are related, they are not synonymous.

An intuitively appealing approach to assessing the amount of inheritance in an occupational structure would be to correlate the origin with destination position. If the correlation is unity, then it may be inferred that there is complete determination between origins and destinations, in the sense that there is no uncertainity of destination, if the origin is known. This strategy is often implemented by mapping the origin and destination occupational positions into a conceptually continuous variable such as the North-Hatt occupational ratings, or Duncan's socio-economic index.[1] This approach may be justified in many circumstances, but it does demand that assumptions concerning the nature of the phenomenon, as well as its metric, be built into the analysis. Specifically, the adoption of such a strategy requires that we recast the mobility problem from investigating the movement of manpower in the occupational structure to mobility within a status hierarchy. Although the work of Klatzky and Hodge (1971), Duncan-Jones (1972), as well as the original treatise of Duncan (1961), demonstrate a close correspondence between occupational membership and socioeconomic ranking, it remains that this relationship is not exact, and conceptually the two dimensions are not completely interchangeable. It is possible, for example, to change occupations without altering the worker's socioeconomic status, and for many sociological problems, the movement of labor among occupations may be more interesting than changes in worker's status.

In this paper we will suggest a procedure by which it is possible to determine the degree of inheritance in an occupational structure without resorting to translating the origin and destination positions into prestige or status scale scores, therefore avoiding becoming encumbered with perhaps unwanted, and possibly unwarranted, conceptual as well as mathematical assumptions.

Development of the Model

The most common means by which to assess the amount of mobility within a given area is the use of the transition matrix. The matrix fully describes the amount of mobility that takes place within the transition interval, and it represents the process by which the distribution of the labor force is created in one time period as a function of the distribution of labor in the previous time. The transition matrix, then, summarizes much information concerning the amounts and patterns of mobility during a particular time interval.

Let O^t be a vector of $\underline{k}$ proportions whose typical element $\{O_j^t\}$ is the proportion of the labor force in the $\underline{j}^{th}$ occupational group at time $\underline{t}$.

By definition

$$\sum_{j=1}^{k} O_j^t = 1 \qquad (1)$$

these elements sum to unity. Similarily, let O^{t-1} be a vector of k proportions whose typical element $\{O_i^{t-1}\}$ is the proportion of the labor force in the i^{th} occupational group at some earlier point in time, $t-1$. Again,

$$\sum_{i=1}^{k} O_i^{t-1} = 1. \qquad (2)$$

Vectors O^{t-1} and O^t are related by a square $k \times k$ transition matrix, M:

$$O^t = M \cdot O^{t-1}. \qquad (3)$$

The typical element of this matrix $\{M_{ij}\}$ is a transition probability, or outflow coefficient, and it is interpreted as the conditional probability that a worker is in occupational group j at time t given that the worker was in occupational group i at the earlier time, $t-1$:

$$M_{ij} = \Pr\{O_j^t | O_i^{t-1}\}. \qquad (4)$$

From this definition it follows that the elements in the i^{th} row of the transition matrix will sum to unity

$$\sum_{j=1}^{k} M_{ij} = 1. \qquad (5)$$

In the case of perfect retention in the i^{th} occupational group from time $t-1$ to t, the i^{th} diagonal element of M will be unity. Likewise, complete mobility out of the i^{th} group would be indicated by a null diagonal element. The magnitude of the diagonal elements of the transition matrix provide a ready index to the amount of <u>cell-specific</u> mobility, but they do not indicate the degree of inheritance for the whole matrix, although they are often used in this regard. A transition matrix which demonstrates a great deal of mobility, diagonal elements near null, may still be quite determinant. In sum, little information is provided by the diagonal elements of the transition matrix for assessing the degree to which origin positions determine the destinations for the occupational structure as a whole.

One approach to investigating the question of inheritance is to correlate the two marginal distributions of the transition matrix, O^t and O^{t-1}. In this regard, a canonical correlational analysis would seem advisable. The suggested procedure is to first define two sets of binary variables:

$$D_{ij}^{t-1} = \begin{cases} 1 & \text{If the } i^{th} \text{ worker is in the } j^{th} \text{ of } k \text{ occupational groups in time } t-1 \\ 0 & \text{Otherwise} \end{cases} \qquad (6)$$

$$D_{ij}^{t} = \begin{cases} 1 & \text{If the } i^{th} \text{ worker is in the } j^{th} \text{ of } k \text{ occupational groups in time } t \\ 0 & \text{Otherwise.} \end{cases} \qquad (7)$$

The first set of k binary variables defines the i^{th} worker's occupational membership in time t-1, while the second defines the position at the end of the transition interval, time t. Since the i^{th} worker must be located in one of the k occupational groups, by definition there are only k-1 linearly

$$\sum_{j=1}^{k} D_{ij}^{t} = \sum_{j=1}^{k} D_{ij}^{t-1} = 1 \qquad (8)$$

independent elements in the vectors D_j^t and D_j^{t-1}.[2] Now consider two unmeasured variables S and F, each of which is a weighted linear combination of k-1 of the binary variables in each set:

$$S_i = \sum_{j=1}^{k} a_j D_{ij}^{t-1} \quad (9)$$

$$F_i = \sum_{j=1}^{k} b_j D_{ij}^{t} \quad (10)$$

The purpose of canonical analysis is to choose the weights, $\underline{a_j}$'s and $\underline{b_j}$'s, such that the product-moment correlation between variables $\underline{S}$ and $\underline{F}$ is maximized. There is, however, no unique solution to this problem since it is possible to multiply each of the chosen canonical weights by a nonzero constant and acquire a new set of scores which will also maximize the correlation (Van de Geer, 1971). Interest here is, however, upon the canonical correlations per se, and not on the canonical coefficient so the problem presented above is of little import at this juncture.[3]

In canonical analysis it is possible to obtain as many orthogonal canonical solutions to equations (9) and (10) as there are linearly independent variables in the smallest set being analyzed. In our case, the number of independent binary variables is the same in each set, and is equal to the k-1 occupational groups. It is possible, therefore, to generate k-1 canonical variates and canonical correlations.[4] In other words, there are k-1 possible solutions to equations (9) and (10). This multiplicity of possible answers raises the question as to which solution is to be selected as being the correlation between vectors O^t and O^{t-1}. The most commonly adopted tactic is to report the canonical correlation associated with the first canonical variate as the correlation between the two sets of binary variables (cf. Klatzky and Hodge, 1971; Duncan-Jones, 1972). There are, however, alternatives to this practice and these will be discussed next.

In a recent investigation Srikantan (1970) compared three measures suitable for assessing the amount of association between two sets of binary variables: (1) the mean square canonical correlation (MSCC); (2) the geometric mean canonical correlation (GMCC); and (3) the squared vector multiple correlation (SVMC). To this list we may add a fourth: the maximum canonical correlation squared (MCCS). These four alternatives are defined as:

$$\mathrm{MSCC} = \frac{1}{k-1} \sum_{j=1}^{k-1} r_j^2 \tag{11}$$

$$\mathrm{GMCC} = \left(\prod_{j=1}^{k-1} r_j \right)^{1/(k-1)} \tag{12}$$

$$\mathrm{SVMC} = 1 - \left(\prod_{j=1}^{k-1} (1 - r_j^2) \right) \tag{13}$$

$$\mathrm{MCCS} = \mathrm{Max}\{r_j^2\} \tag{14}$$

where r_j is the j^{th} of the $k-1$ canonical correlations. Srikantan concludes that the MSCC is to be prefered over the GMCC and SVMC on two grounds: (1) MSCC will be unity if, and only if, the two sets of binary variables are completely independent; and (2) MSCC is directly related to Cramer's V thus it is possible to derive a sampling distribution and test statistic for MSCC. Although Srikantan did not entertain the fourth alternative, MCCS would seem to be a likely candidate. It lacks, however, efficiency by not fully exploiting all the available information concerning the relationship between the two sets of variables; that is, MCCS ignores the remaining $k-2$ solutions to the canonical problem. Of the four alternatives presented above, the MSCC would appear to be the best choice of an index of association between the vectors 0^t and 0^{t-1}.

MSCC can be interpreted as the proportion of the variance in the distribution of the labor force at time t that is explained by the distribution of the labor force at an earlier point in time, t-1. If during a given transition period there is perfect inheritance, the MSCC will be unity. If, however, the movement is essentially unpatterned, MSCC will be null. MSCC thus provides a clear indicator to the extent of inheritance in an occupational structure for a given time interval. Obviously, (1-MSCC) can be used as a measure of the amount of randomness, or non-inheritance within the system between the two points in time.

An Application: Economic Diversification and Occupational Mobility

While space does not permit a full discussion of the relevant theoretical framework, let it suffice to indicate that there is a great deal of literature, both theoretical and empirical, relating economic change and occupational mobility. We will utilize the canonical approach advocated previously to investigate the effects of economic diversification upon short-term intra-generational occupational movement.

It is useful to view occupational mobility within the context of a supply-demand model. The position taken here is that the economic diversification stimulated by industrial development influences the demand aspect of the model and that mobility is one process by which this demand may be satisfied. It has been noted that economic diversification results in a more segmentalized production system and differentiated occupational structure, which requires an increasingly specialized and skilled labor force. The diversification from an economy founded primarily on non-industrial activities to one based more on the factory results in the creation of occupational vacancies. One mechanism for filling these vacancies is occupational mobility. We will investigate one such case of economic diversification, and attempt to assess the impact of development upon the occupational structure.

In 1966 Jones-Laughlin Steel Corporation (J&L) began construction on a highly automated production complex in Putnam County, Illinois. Putnam County is located about 100 miles west of the city of Chicago and, until the introduction of the J&L facility, it was primarily agricultural in nature. As of spring, 1972, J&L had a labor force of about 1000 workers; of these, about two-thirds held occupations that would be considered "blue collar". Soon after J&L announced their plans for the construction of the plant, a team of social scientists at the University of Illinois started an investigation of the impact of the J&L complex on the local communities.[5] An area-probability sample survey of heads of households in the Putnam County area was conducted in 1966. These respondents were re-interviewed the following year, and again four years later. In 1971 a new area-probability survey was completed. Data from this second cross-sectional survey will be utilized in this analysis. The 1971 survey of the developing regions yielded a total of 1166 observations, but for this analysis certain subgroups will be excluded from the basic data set, -specifically, all females, all non-white males, and all males who have not been residing continuously in the study area since 1965.[6] These restrictions reduced the data set to 692 observations.

Detailed yearly work histories covering the period 1966 through 1971 were obtained from each respondent. These histories were coded originally into three-digit U.S. Census codes for occupations and industries. For this analysis, these were recoded into seven major occupational groups.[7] For each of the respondents, we have, then, a yearly history which locates their position in the occupational structure vis-a-vis the seven occupational categories. These yearly histories were translated into six sets, of seven binary variables each, according to the conventions set forth in equations (6) and (7). Each binary variable indicates the head of household's membership in one of the occupational groups for each year, 1966 through 1971. These sets of binary variables provide the basic data input for the canonical analysis.[8] The results of the investigation are presented next.

Table 1 presents the MSCC values for the inter-annual, and lagged, occupational transition matrices. The first column of Table 1 gives the association between the inter-annual, one-year lag matrices. There is an

Table 1 About Here

increasing correlation between the annual transitions from the 1966-1967 matrix to the 1969-1970 transition, although the difference in MSCC for the 1969-1970 and 1968-1969 matrices is sufficiently small to be a function of sampling error. The largest increase in the degree of transition predictability occurred between the 1966-1967 and 1967-1968 matrices. The 1967-1968 transition is about 5% more determinant than the 1966-1967 matrix. This is interpreted as reflecting the impact of the opening of the J&L facility in 1967. It would appear that the introduction of the plant increased the predictability of occupational transitions.

Interestingly, the association in the 1970-1971 period is considerably less than the MSCC for the 1969-1970 transition. This would seem to suggest the occurance of some event during 1970 which resulted in a marked increase in unpatterned occupational movement between 1970 and 1971. At this point we are unable to reach a satisfactory explanation of the factors which could account for the decreased predictability in the 1970-1971 period.

The first row in the table gives the MSCC values between the 1966 occupations and each subsequent annual occupational distribution. One clear pattern emerges: starting with the 1966-1967 transition, there is an approximate 7% decline in MSCC as the lag increases. Thus, the MSCC for the 1966-1968 transition is about 7% less than the MSCC for the 1966-1967 period; likewise, the 1966-1969 MSCC is about 7% less than the 1966-68 transition, and so forth. This indicates that the changes that have evolved in the 1966 occupations occurred in a relatively stable fashion.

The elements along the diagonal of Table 1 are the mean square canonical correlations between the 1971 occupations and each prior annual distribution. There is a declining amount of association between the 1971 occupations and each of the others. The greatest difference in MSCC occurs between the 1967-1971 MSCC and the 1966-1971 MSCC. There is about 12% more predictability in the 1967-1971 transition than in the 1966-1971 period. This reinforces the view that the major influence of the J&L complex took place during 1967, the first year of production.

In summary, the canonical analysis of the occupational transition matrices revealed that the economic diversification, stimulated by the introduction of the steel mill, appears to have resulted in a substantial increase in the stability of occupational transitions from the opening of the plant in 1967 through 1970. Yet some unidentified factor, or factors, seems to have altered this stability during the period 1970-1971.

Summary

We have outlined a procedure for determining the amount of predictability, or inheritance, in an occupational structure that avoids translating the origin and destination positions into scale scores, thus altering the conceptual context from occupational mobility to movement through a status hierarchy. The recommended technique involves recoding the occupational positions into two sets of binary variables, then obtaining the mean square canonical correlation (MSCC) between the two sets. The MSCC provides a convenient measure to which the origin position determines the destination. MSCC will attain unity if there is complete inheritance in the occupational system, and will reach zero in the absence of any predictability.

Although we have framed this paper in terms of occupational mobility, the procedure is readily generalizable to many different types of problems that are amenable to matrix formulation.

NOTES

[1]See Robinson, et al. (1969) for a detailed comparison of various occupational indices and scales.

[2]See Suits (1957) and Melichar (1965) for a discussion of collinearity and binary variables.

[3]See Hodge and Klatzky (1971) for a discussion of this point. Klatzky and Hodge (1971) in their re-analysis of the Blau-Duncan OCG data do concentrate on interpreting the canonical coefficients. Duncan-Jones (1972) uses the canonical weights to scale occupational groups.

[4]Only the first canonical variate will generate the maximum product-moment correlation between vectors O^t and O^{t-1}.

[5]See Summers, et al. (1969) for a discussion of the research design and description of the study area.

[6]The rationale for these exclusions is outlined in Beck (1972).

[7]The seven occupational categories were: (1) professional and technical workers; (2) farmers, farm managers, and farm laborers; (3) managers, officials, and proprietors; (4) clerical and sales workers; (5) craftsmen, foremen, and operatives; (6) private household workers, other service workers, and laborers; (7) those not active in the civilian labor force.

[8]Because of the collinearity problem each set contained only six binary variables. For the purposes of this analysis, it is immaterial which binary variable is deleted.

Table 1

MSCC Values for Occupational Transition Matrices

T	1967	1968	1969	1970	1971
1966	.8234	.7527	.6857	.6171	.5501
1967	.8708	.7918	.7188	.6697	
1968	.8924	.8074	.7073		
1969	.9042	.7736			
1970	.8521				

REFERENCES

Bartholomew, David J.

1967 Stochastic Models for Social Processes. London: John Wiley & Sons.

Beck, E.M.

1972 A Study of Rural Industrial Development and Occupational Mobility. Unpublished Ph.D. thesis, University of Tennessee.

Blau, Peter M. and Otis Dudley Duncan

1967 The American Occupational Structure. New York: John Wiley & Sons.

Blumen, Isadore, Marvin Kogan, and Philip J. McCarthy

1955 The Industrial Mobility of Labor as a Probability Process. Ithaca, New York: Cornell University Press.

Carter, T. Michael and Nancy C. Carter

1971 "Status Attainment vs. Social Mobility: Alternative or Complementary Modes of Stratification Research." LSU Journal of Sociology 1 (October): 12-33.

Duncan, Otis Dudley

1961 "A Socioeconomic Index for all Occupations." Pp. 109-138 in Albert J. Reiss, Jr., et al., Occupations and Social Status. New York: The Free Press.

Duncan-Jones, P.

1972 "Social Mobility, Canonical Scoring and Occupational Classification." Pp. 191-210 in Keith Hope (ed.), The Analysis of Social Mobility. Oxford: Clarendon Press.

Glass, David V.

1954 *Social Mobility in Britain.* London: Routledge and Kegan Paul.

Hodge, Robert W. and Sheila R. Klatzky

1971 *Canonical Analysis of Discrete Variables.* Working Paper 71-1, Madison, Wisconsin: Center for Demography and Ecology, University of Wisconsin.

Klatzky, Sheila R. and Robert W. Hodge

1971 "A Canonical Correlation Analysis of Occupational Mobility." *Journal of the American Statistical Association* 66 (March): 16-22.

Lipset, Seymour Martin and Reinhard Bendix

1959 *Social Mobility in Industrial Society.* Berkeley, Calif.: University of California Press.

Melichar, Emanuel

1965 "Least-Squares Analysis of Economic Survey Data." Paper presented at the Economic Statistics Section, American Statistical Association Annual Meetings, Philadelphia, Penn.

Robinson, John P., *et al.*

1969 *Measures of Occupational Attitudes and Occupational Characteristics.* Ann Arbor, Mich.: Institute for Social Research, University of Michigan.

Rogoff, Natalie

1953 *Recent Trends in Occupational Mobility.* Glencoe, Ill.: The Free Press.

Srikantan, K.S.

1970 "Canonical Association Between Nominal Measurements." Journal of the American Statistical Association 65 (March): 284-292.

Suits, Daniel B.

1957 "Use of Dummy Variables in Regression Equations." Journal of the American Statistical Association 52 (December): 548-551.

Summers, Gene F., et al.

1969 Before Industrialization: A Social System Base Study of a Rural Area. Urbana, Ill.: Illinois Agriculture Experiment Station, University of Illinois.

Van de Geer, John P.

1971 Introduction to Multivariate Analysis for the Social Sciences. San Francisco, Calif.: W. H. Freeman and Co.