

A Canonical Approach to Assessing Occupational Mobility Matrices¹

E. M. BECK

University of Michigan

This paper develops a summary measure of occupational persistence which does not require translating the occupational categories into status scale scores. The measure is based on the mean of the squared canonical correlations extracted from a general mobility matrix. The procedure is employed to assess the impact of rapid regional economic diversification on the movement of labor within the occupational structure.

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 evaluation 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. Hence, the actual rate of mobility is compared to the rate anticipated if the origin positions had no effects on determining the destinations (Bartholomew, 1967). Departures away from independence are then interpreted as the effects of "inheritance."² Occupational systems which demonstrate considerable mobility are then held to manifest minimal rigidity and to be relatively open.

An intuitively appealing approach to assessing the amount of inheritance in an occupational structure would be to correlate the origin with destination positions. If the correlation is unity, then it may be inferred that there is

¹An earlier version of this paper was presented at the annual meetings of the Pacific Sociological Association, May, 1973. Support for the research was provided by the College of Agriculture and Life Sciences at the University of Wisconsin, Madison, and Public Health Service Research Grant MH-19689 from the National Institute of Mental Health to Gene F. Summers and John P. Clark. The comments of David L. Featherman, Hanan C. Selvin, and Gene F. Summers proved insightful and beneficial, although they were not always adopted.

²Inheritance is used throughout the paper to refer to occupational persistence—the nonmovement, either inter- or intragenerationally, within the occupational structure.

complete determination between origins and destinations, in the sense that there is no uncertainty of destination given that the origin is known. This tactic 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 (1961) socioeconomic index. 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 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 rank, 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 workers' status.

In this paper we will suggest a procedure by which it is possible to summarize 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, or possibly unwarranted, conceptual as well as mathematical assumptions. The task at hand is to devise some indicator of inheritance based on the correspondence between the origin and destination occupational positions. Following Srikantan (1970), we hold that this measure should embody three properties:

1. the measure should be zero if, and only if, the origins and destinations are stochastically independent,
2. it should attain unity if, and only if, there is no uncertainty of destination if the origin is known,
3. the measure should incorporate all available information that can be derived from the mobility matrix which links the origins and destinations.

One strategy, which has been suggested by others (cf. Klatzky and Hodge, 1971; Duncan-Jones, 1972; Hodge and Klatzky, 1971), involves formulating the problem in terms of canonical analysis. This approach will be used here as a tactic for assessing the degree of nonmobility in occupational structures. Anticipating the formal development of the measure, we will show that it conforms to the three desired characteristics outlined above.

DEVELOPMENT OF THE MODEL³

Let \mathbf{O} be a vector of k origin occupational positions whose typical element $[o_i]$ is the number of workers in the i th of k occupational categories at a given point in time. The sum of these elements is the total number of

$$\sum_{i=1}^k o_i = N$$

workers in the labor force, N .

Similarly, let \mathbf{D} be a vector of k destination positions whose typical element $[d_j]$ is the number of workers in the j th occupational position at a later time. Hence,

$$\sum_{j=1}^k d_j = N.$$

Vectors \mathbf{O} and \mathbf{D} are the row and column marginal distributions of a square $k \times k$ mobility matrix \mathbf{M} whose typical element $[m_{ij}]$ is the number of workers who moved from the i th origin position to the j th destination during the transition interval. Given these definitions, the following relations hold for vectors \mathbf{O} and \mathbf{D} , and matrix \mathbf{M} .⁴

$$\sum_{j=1}^k m_{ij} = o_i,$$

$$\sum_{i=1}^k m_{ij} = d_j,$$

$$\sum_{i=1}^k \sum_{j=1}^k m_{ij} = \sum_{i=1}^k o_i = \sum_{j=1}^k d_j = N.$$

Now consider a vector of k weights, each being associated with one of the origin positions, hence

$$\xi_i = \text{the weight assigned to the } i\text{th origin.}$$

Likewise, consider another vector of k weights associated with each element of the destination vector:

$$\eta_j = \text{the weight assigned to the } j\text{th destination.}$$

³Much of this discussion is an abbreviated version of more technical literature. Detailed elaborations and proofs can be found in Srikantan (1970), Lancaster (1969), Van de Geer (1971), Williams (1952), and Marriott (1952).

⁴Let \mathbf{P} be a $k \times k$ standardized mobility matrix where the typical element $[p_{ij}]$ is the conditional probability that a worker is in the j th destination given that the worker started in the i th origin. Conventionally, \mathbf{P} is known as a transition matrix, and the p_{ij} s are transition probabilities, or outflow coefficients. Vectors \mathbf{O} and \mathbf{D} are related by this transition matrix, hence

$$\mathbf{D} = \mathbf{P}' \mathbf{O}.$$

These weights are subject to the constraint

$$\sum_{i=1}^k \xi_i o_i = \sum_{j=1}^k \eta_j d_j = 0.$$

We shall take as a measure of occupational persistence the correlation between the weighted origin vector $[\xi \mathbf{O}]$ and the weighted destination vector $[\eta \mathbf{D}]$.

At this point it is necessary to define two new matrices: let \mathbf{X} be a $k \times k$ diagonal matrix whose typical element $[x_{ii}]$ is the number of workers in the i th origin position; that is, $x_{ii} = o_i$. Likewise, let \mathbf{Y} be a $k \times k$ diagonal matrix whose typical element $[y_{jj}]$ is the number of workers in the j th destination; that is, $y_{jj} = d_j$. Now, the ξ_i 's and η_j 's can be obtained from eigenvectors extracted from matrix \mathbf{T}

$$\mathbf{T} = \mathbf{X}^{-1} \mathbf{M} \mathbf{Y}^{-1} \mathbf{M}', \quad (1)$$

where \mathbf{M} is the mobility matrix, and matrices \mathbf{X} and \mathbf{Y} are defined above. Since we are assuming that \mathbf{T} is full rank, it is possible to obtain k nonzero eigenvectors. The first eigenvector is a trivial solution and can be dismissed from further consideration (Williams, 1952; Srikantan, 1970). The remaining $k-1$ eigenvectors represent m sets of unstandardized canonical weights.⁵ In other words, there are m sets of nontrivial solutions for the ξ_i 's and η_j 's; hence, $\xi^{(m)}$ and $\eta^{(m)}$ are the m th set of canonical scores.

Associated with each eigenvector is an eigenvalue which can be shown to be the squared canonical correlation between $[\xi^{(m)} \mathbf{O}]$ and $[\eta^{(m)} \mathbf{D}]$. Thus, ρ_m is the correlation between the weighted origin vector and weighted destination vector that is associated with the m th canonical solution. This correlation is obtained by taking the square root of the m th eigenvalue.

In a recent investigation Srikantan (1970) compares three measures of association based on these m canonical correlations: the Mean Square Canonical Correlation (MSCC); the Geometric Mean Canonical Correlation (GMCC); and the Squared Vector Multiple Correlation (SVMC). These three alternatives are defined as:

$$\text{MSCC} = \frac{1}{k-1} \sum_{m=1}^k \rho_m^2, \quad (2)$$

$$\text{GMCC} = \left(\prod_{m=1}^k \rho_m^2 \right)^{1/(k-1)}, \quad (3)$$

$$\text{SVMC} = 1 - \prod_{m=1}^k (1 - \rho_m^2), \quad (4)$$

where ρ_m is the m th canonical correlation.

⁵The procedure for standardizing the eigenvectors is discussed by Duncan-Jones (1972) and Srikantan (1970).

Srikantan concludes that the MSCC is to be preferred over the GMCC and SVMC on two grounds: the MSCC will be unity if the origins and destinations are completely dependent, and it will be zero if they are stochastically independent; secondly, the sample estimator of MSCC has a known sampling distribution, hence a statistical test of significance is available.

The MSCC, which we'll denote as $\bar{\rho}^2$, can be interpreted as the proportion of the variance in the weighted destination vector that is explained by the weighted origin vector. If, over a given period, there is complete occupational stability (a diagonal mobility matrix), $\bar{\rho}^2$ will be unity. If, however, the mobility matrix is random, $\bar{\rho}^2$ will be null. In sum,

$$\begin{aligned}\bar{\rho}^2 &= 1 && \text{iff } m_{ij} = 0 && \text{for all } i \neq j, \\ \bar{\rho}^2 &= 0 && \text{iff } m_{ij} = o_i d_j / N && \text{for all } i \text{ and } j.\end{aligned}$$

It is our position, then, that $\bar{\rho}^2$ provides a clearly interpretable summary measure of inheritance in mobility tables because it reflects the degree to which information pertaining to the origin position will increase our knowledge of the destination position. It varies, $0 \leq \bar{\rho}^2 \leq 1$, and will attain these limits only in the case of complete stochastic independence, or complete dependence. Lastly, it is efficient in the sense that $\bar{\rho}^2$ incorporates all information from the m nontrivial canonical solutions to the mobility matrix.

If it is desired to obtain all the canonical correlations, we may extract the nontrivial eigenvalues, and their associated eigenvectors, from matrix T , as defined in Eq. (1), and $\bar{\rho}^2$ can be computed from Eq. (2). If there is little interest in acquiring the full array of canonical correlations, $\bar{\rho}^2$ can be computed in a more straightforward fashion (Srikantan, 1970):

$$\bar{\rho}^2 = [(\text{Trace of } T) - 1] / (k-1). \quad (5)$$

Or, it can be derived directly from the mobility matrix (Lancaster, 1969; Srikantan, 1970):

$$\bar{\rho}^2 = \frac{\left[\sum_{i=1}^k \sum_{j=1}^k (m_{ij}^2 / o_i d_j) \right] - 1}{k-1}. \quad (6)$$

To this point we have only considered $\bar{\rho}^2$, the *population* mean square canonical correlation. \bar{r}^2 , a consistent estimator of $\bar{\rho}^2$, can be obtained by replacing population vectors O and D , and matrix M , with their sample analogs. \bar{r}^2 is then computed using either Eq. (2), (5), or (6). If a statistical test of significance is desired, it has been shown that under the null hypothesis the quantity $\bar{r}^2 (nk - n)$ is distributed as χ^2 with $(k-1)^2$ degrees of freedom, where n is the sample size (Srikantan, 1970; Lancaster, 1969).⁶

⁶Marriott (1952) and Lancaster (1969) give detailed discussions of tests of statistical significance in canonical analysis.

AN APPLICATION: ECONOMIC DIVERSIFICATION AND OCCUPATIONAL STABILITY

Though space does not permit a full discussion of the relevant theoretical framework, there is a great deal of literature, both theoretical and empirical, relating economic growth and occupational mobility. We will utilize the canonical approach to investigate the effects of rapid economic diversification upon the stability of labor in the occupational structure.

It is useful to view occupational movement within the context of a supply-demand model. The economic diversification that may be stimulated by industrial growth influences the demand aspect of the model, and occupational mobility is a process by which this demand may be satisfied. As Kerr *et al.* (1960) and Levitan and Sheppard (1963) have noted, industrial development yields a more segmentalized production system and more highly differentiated occupational structure. The diversification of an economy founded primarily on agricultural pursuits to one based more on the factory results in the withering away of some old occupations and the growth of new ones (Levitan and Sheppard, 1963; Smelser, 1963; Smelser and Lipset, 1966; Hoselitz, 1955). By creating new occupations, or occupational categories, economic diversification results in demands to fill these occupational vacancies. This demand can be supplied in three ways: (1) the migration of workers into the developing region; (2) occupational mobility of workers into the vacancies; or (3) differential fertility among different occupational groups. Over a short period of time, the third alternative is not plausible, thus to supply the labor demands caused by economic diversification there must be either in-migration or occupational mobility, or some combination of both.

The conditions which promote mobility versus in-migration, however, are not altogether clear. One of the factors which would seem to determine whether the occupational demands will be met by mobility or migration is the degree of consistency between the skill levels needed by the industrial inputs and the availability of these skills in the local market. The skill demand will be largely determined by the size, and technological sophistication, of the force stimulating the economic diversification. It would be reasonable to anticipate that there will be increased in-migration in situations where the skills demanded cannot be supplied by the local labor force. If, however, we control the amount of in-migration, it would be reasonable to expect that economic diversification should stimulate movement within the occupational structure by providing some alternative avenues of employment in a fundamentally nonindustrial economy (Kerr *et al.*, 1960; Moore, 1966; Smelser and Lipset, 1966). The direction of this movement, and even its amount, are difficult to forecast with any degree of accuracy (Duncan, 1966). Further, it might be reasoned that the diversification of an agrarian economy may also increase the regularity of work and further stabilize the labor force. By

creating nonagriculture-related occupations, which are less sensitive to variation in demand, development may, in the long run, routinize the organization of labor in a more stable configuration.

It might be anticipated, then, that rapid economic diversification may initially stimulate the movement of labor within the occupational structure, but after some passage of time the labor force would tend to stabilize. It is difficult to predict, however, how rapidly the effects of industrial growth may dampen. We will investigate one such case of diversification and attempt to assess its impact 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, the region was primarily agricultural in nature. The cold-rolling component of J&L's works has been in operation since December, 1967; when production began the mill's labor force numbered about 700. As of Spring, 1972, J&L had a labor force of about 1000 workers of which most—about 2/3—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.⁷ 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 4 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 yielded a total of 1166 observations.

Ten-year annual residential histories were obtained from each respondent. Using information from these residential data, respondents who have not resided continuously in the Putnam County area since the introduction of the J&L plant were excluded. This effectively limits the analysis to the native labor force and controls on the effects of migration into, as well as out of, the region. Other restrictions were also placed on the data. All female heads of households as well as nonwhite heads were excluded for this investigation. Since rural female heads of households are, by and large, elderly who are inactive in the labor force, they were eliminated from the mobility analysis. The nonwhite heads were deleted because of their extremely small number in the population being sampled. These restrictions reduced the original 1971 data set from 1166 observations to 692.

Detailed annual work histories covering the period 1966-1971 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: (1) professional and

⁷See Summers *et al.* (1969) and Beck (1972) for a discussion of the research design and description of the study area.

TABLE 1

Intercorrelations Among Occupational Distributions, 1966-1971

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

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 and other service workers, and laborers; and (7) those not active in the civilian labor force. For each of the respondents we have, then, a yearly history which locates their position in the occupational structure *vis-à-vis* the seven occupational categories. Using these basic data, the consistent estimators of the population mean square canonical correlations were computed.

The mean square canonical correlations for the interannual and lagged occupational distributions are given in Table 1. Using the χ^2 test described previously, we found that all the correlations are significant at the .01 level. The elements along the main diagonal of this tabulation are the interannual occupational correlations, e.g., 1966-1967, 1967-1968, etc. Starting with the 1966-1967 correlation, the interannual association increases monotonically from .8234 to .9042 for the 1969-1970 period. This reveals that there has been a decreasing amount of interannual occupational mobility from 1966 to 1970. The largest decrease occurred between the 1966-1967 and 1967-1968 periods: there was about 5% more occupational stability in the 1967-1968 transition ($\bar{r}^2 = .8708$) than in the 1966-1967 period ($\bar{r}^2 = .8234$). This may be interpreted as reflecting the impact of the opening of the J&L mill in 1967. It would appear that the introduction of the plant has resulted in a greater interannual correlation among occupations, hence increased stability in the labor force. Interestingly, the MSCC for the 1970-1971 period ($\bar{r}^2 = .8521$) is considerably less than for the 1969-1970 transition ($\bar{r}^2 = .9042$). This would suggest the occurrence of some event during 1970 which resulted in a marked increase in occupational mobility. At this time we are unable to reach a satisfactory explanation of the factors which could account for this increased movement in the labor force during 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 yearly lag increases. Thus the association between the 1966-1968 occupations ($\bar{r}^2 = .7527$) is about 7% less than the

MSCC for the 1966-1967 transition ($\bar{r}^2 = .8234$); likewise, the 1966-1969 MSCC ($\bar{r}^2 = .6857$) is about 7% less than the 1966-1968 MSCC, and so forth. This indicates that the changes that have evolved in the 1966 occupational distribution occurred in a relatively stable fashion. There is an approximate 7% annual decay in the correlation between the 1966 occupations and each subsequent annual distribution. The last column of Table 1 presents the correlations between the 1971 occupational distribution and each *prior* annual distribution. There is a declining amount of association between the 1971 occupations and each of the previous others. The greatest change occurs between the 1967-1971 and the 1966-1971 periods: there is about 12% more correlation between the 1967-1971 occupations ($\bar{r}^2 = .6697$) than between the 1966-1971 occupations ($\bar{r}^2 = .5501$). This reinforces the view that the major influence of the J&L complex took place during 1967, the first year of plant production.

In summary, the canonical analysis of the occupational mobility 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 start of production in 1967 through 1970, and the greatest impact of the plant appears to have been upon its initial opening. Yet some unidentified factor, or factors, seems to have altered this stability during the period 1970-1971.

CONCLUSIONS

This paper has described a method for assessing the amount of movement within a general mobility matrix. The procedure is not intended to replace detailed inspection and interpretation of mobility tables, rather its function is to supplement such analysis. As a descriptive tool, it provides a convenient summary measure of the association between occupational origins and destinations which does not require that the occupations be rescored into status scale scores. The basic strategy is to cast the problem in terms of canonical analysis and to formulate the measure as the mean of the squared canonical correlations.

Although the discussion here has been framed in terms of a symmetric mobility matrix, the canonical method can be generalized to other problems which are amenable to matrix representation. Further, with a very slight modification of the statistical model, the procedure can be applied to asymmetrical $r \times c$ matrices. In such cases there will be either $(r-1)$ or $(c-1)$, depending on which is smaller, nontrivial eigenvalues; hence the MSCC will be based on either $(r-1)$ or $(c-1)$ canonical correlations.

REFERENCES

- Bartholomew, David J. (1967), *Stochastic Models for Social Processes*. Wiley, London.
- Beck, E. M. (1972), *A Study of Rural Industrial Development and Occupational Mobility*. Unpublished Ph.D. thesis, University of Tennessee, Knoxville.
- Blumen, Isadore, Kogan, Marvin, and McCarthy, Philip J. (1955), *The Industrial Mobility of Labor as a Probability Process*. Cornell University Press, Ithaca, New York.
- Duncan, Otis Dudley (1961), "A socioeconomic index for all occupations." *in Occupations and Social Status*, (Albert J. Reiss with collaborators, Eds.), pp. 109-138. The Free Press, New York.
- Duncan, Otis Dudley (1966), "Methodological issues in the analysis of social mobility." *in Social Structure and Mobility in Economic Development*, (Neil J. Smelser and Seymour Martin Lipset, Eds.), pp. 51-97. Aldine, Chicago.
- Duncan-Jones, P. (1972), "Social mobility, canonical scoring and occupational classification." *in The Analysis of Social Mobility* (Keith Hope, Ed.), pp. 191-210. Clarendon Press, Oxford.
- Glass, David V. (1954), *Social Mobility in Britain*. Routledge and Kegan Paul, London.
- Hodge, Robert W., and Klatzky, Sheila R. (1971), *Canonical Analysis of Discrete Variables*. Working Paper 71-1, Center for Demography and Ecology, University of Wisconsin, Madison.
- Hoselitz, Bert F. (1955), "The city, the factory, and economic growth." *American Economic Review* 45, 166-184.
- Kerr, Clark, Dunlop, John T., Harbison, Frederick, and Myers, Charles A. (1960), *Industrialism and Industrial Man*. Harvard University Press, Cambridge.
- Klatzky, Sheila R. and Hodge, Robert W. (1971), "A canonical correlational analysis of occupational mobility." *Journal of the American Statistical Association* 66, 16-22.
- Lancaster, H. O. (1969), *The Chi-Squared Distribution*. Wiley, New York.
- Levitan, Sar A. and Sheppard, Harold L. (1963), "Technological change and the community." *in Adjusting to Technological Change* (Gerald G. Sommers, Edward L. Cushman, and Nat Weinberg, Eds.), pp. 159-189. Harper & Row, New York.
- Lipset, Seymour Martin, and Bendix, Reinhard (1959), *Social Mobility in Industrial Society*. University of California Press, Berkeley.
- Marriott, F. H. C. (1952), "Tests of significance in canonical analysis." *Biometrika* 39, 58-64.
- Moore, Wilbert E. (1966), "Changes in occupational structures." *in Social Structure and Mobility in Economic Development* (Neil J. Smelser and Seymour Martin Lipset, Eds.), pp. 194-212. Aldine, Chicago.
- Rogoff, Natalie (1953), *Recent Trends in Occupational Mobility*. The Free Press, Glencoe, Illinois.
- Smelser, Neil J. (1963), "Mechanisms of change and adjustment to change." *in Industrialization and Society* (Bert F. Hoselitz and Wilbert E. Moore, Eds.), pp. 32-54. UNESCO, Paris.
- Smelser, Neil J. and Lipset, S. M. (1966), "Social structure, mobility and development." *in Social Structure in Economic Development* (Neil J. Smelser and S. M. Lipset, Eds.), pp. 1-50. Aldine, Chicago.
- Srikantan, K. S. (1970), "Canonical association between nominal measurements." *Journal of the American Statistical Association* 65, 284-292.
- Summers, Gene F., Hough, Richard L., Scott, John T. Jr., and Folse, C. L. (1969), *Before Industrialization: A Social System Base Study of a Rural Area*. Illinois Agricultural Experiment Station, University of Illinois, Urbana.
- Williams, E. J. (1952), "Use of scores for the analysis of association in contingency tables." *Biometrika* 39, 274-289.