HOUSEHOLD LOCATION AND MIGRATION WITHIN THE BOSTON METROPOLITAN REGION

by

PETER MURRAY ALLAMAN

A. B., Otterbein College
(1963)

A. M., Harvard University
(1967)

SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER IN
CITY PLANNING

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1967

Signature of Author . . . .
Department of City and Regional Planning

Certified by . . . .
Thesis Supervisor

Accepted by . . . .
~~Head,~~ Department of City and Regional Planning

ABSTRACT

HOUSEHOLD LOCATION AND MIGRATION
WITHIN THE BOSTON METROPOLITAN REGION

by

PETER MURRAY ALLAMAN

Submitted to the Department of City and Regional Planning on June 19, 1967, in partial fulfillment of the requirement for the degree of Master in City Planning.

The thesis of this study is that migration within a metropolitan area cannot be meaningfully understood except in terms of an explanation of residential location which articulates specific concepts of location theory, as developed in economics, with those of social class, as developed in sociology. Further, that, while location theory, formulated in terms of site rents and transportation costs, provides the overall structuring of a region, social class concepts will account for significant biases not attributable to income differences.

In a preliminary chapter, recent literature touching on both location and migration is reviewed. In subsequent sections, 1960 Census tract data for the Boston region is analysed, primarily by means of canonical correlation and principal components analysis, in an effort to test the thesis.

The principal findings of the study are as follows:

1. The most strongly marked result of the analysis was the emergence of a white-collar/blue-collar clustering of occupational groups, although variations to this pattern were found.

2. A gradient pattern, consisting of a lower amount of residential clustering along class lines for the inner rings of the region, rising to a peak in a middle ring, then falling off toward the outside of the region, was found.

3. Within the framework of canonical correlation, median education, employed as a proxy for "life style" factors, weights more heavily than does median income in indexing the distribution of occupational groups within each Census tract.

Of methodological interest were the following:

4. A principal components analysis of Boston-region census tracts yielded somewhat sharper and more meaningful results than a similar analysis of towns of the Boston area.

5. Canonical correlation appears to have utility in the analysis of ecological data.

Thesis Supervisor: James M. Beshers
Title: Associate Professor of Sociology
Department of City and Regional Planning

ACKNOWLEDGMENTS

The author wishes to thank numerous persons who have contributed to the preparation of this thesis. In particular, he is indebted to Professor James Beshers for his many comments and suggestions, not all of which were heeded. In addition, he is grateful to Miss M. Hyde, Messrs. W. McMains, S. MacIntosh, and D. Griffel for suggestions in the use of their computer systems, to the Harvard and MIT Computation Centers for help with an administrative detail, as well as to the Joint Center for Urban Studies and the MIT Department of City Planning for the use of facilities. To my wife, I owe a special debt of gratitude for her support during the course of this study.

TABLE OF CONTENTS

LIST OF FIGURES AND CHARTS

Chapter I

INTRODUCTION

In order to operate effectively, the urban planner cannot be without a knowledge of urban form or a knowledge of the structural relations which hold the pieces together, or cause them to shift. Since urban studies presently consist of a series of tight compartments, relevant work in a variety of fields may bear on the planner's task. This is particularly true of location and migration studies. Relevant work exists in both economics and sociology. We will consider, in this study, aspects of the formulations in both disiplines, in an endeavor to show how the mechanisms which are proposed in one may bear on the other. In addition, we should also require that the product bear on planning practice.

The theme of this study is that the migration of households within a metropolitan area cannot be meaningfully understood except in terms of an explanation of residential location which articulates specific concepts of location theory, as developed in economics, with those of social class, as developed in sociology. Recent work in both areas is suggestive of the direction in which to proceed. The thesis of this study is that while the location theory formulated in terms of site rents and transportation costs provides the overall structuring of a region, social class considerations will cause significant biases not attributable to income differences. While the household is generally constrained by economic factors, it makes the fine decisions as to location for other reasons.

In the next chapter, we review the several theories and a number of research studies in an effort to relate aspects of each to the other. From economics our principal concerns will be the location theories of Alonzo and Kain. There is a body of research which supports these conceptions, and relevant part of it will be treated, as well. While a great deal of work has not been done in sociology specifically on the topic of the effects of neighborhood proximity, relevant work does exist.

In the next chapter, we apply statistical methodology to the problem. The primary techniques are those of canonical correlation and principal components analysis. The results relate, in part, to other work in the field, although the use of canonical correlation for the analysis of ecological data has not previously appeared.

Chapter II

FACTORS INVOLVED IN LOCATION AND MIGRATION: LITERATURE

Residential Location

The most obvious and most reasonable explanation for patterning arises from a consideration of the linkages of the residential areas to other use areas -- those of employment, shopping, recreation, etc. -- and the complex juggling occurring when numerous groups of households attempt to arrange themselves *vis-a-vis* these destinations. Indeed, one successful urban model[1] needed essentially only the distribution of "basic" employment, trip distribution indices, and structural parameters such as labor force participation rates to explain .62 of the variance in residential population density for sub-areas of the Pittsburgh area.

One useful way of conceptualizing household location is presented by Alonzo[2], whose rent theory expresses the optimum location for a household as that point of tangency of a "bid rent curve" (an indifference curve relating price and distance from core) to a price curve. Several points of clarification need to be made. First of all, the bid-rent curve links the bid price for a *standard unit of land* with distance from the city center; at any point along one curve, the individual household is indifferent to various combinations of land price and transport costs. This implies that those with lower incomes and, other things equal, with less to spend for housing, will tend to have a family of curves with steeper

---

1 I.S. Lowry, *A Model of Metropolis*, Rand (Santa Monica:1964), Table 1, p. 92.

2 W. Alonzo, "Atheory of the Urban Land Market," RSA_P&P:6 (1960), pp. 149-157.

ioration may be a result of greater dispersion, while lowered transport costs and greater average income might tend to flatten all bid-rent curves, in Alonzo's formulation, leading to greater dispersion.

That the bid-rent curve idea is a serviceable concept is particularly evident in the interpretation of known regularities of city growth. Assuming the relatively wealthy to have flatter bid-rent curves, for example, aids in understanding the observations of both Burgess and Hoyt, that the rich tend to always appear on the periphery as the vanguard of growth. Alonzo notes[1] that this process need not necessarily be one of the growing obsolecence of older, more central structures, and the "fleeing" of the more wealthy to the periphery, but may be sustained by the addition of more higher-income families, the improvement of transportation, etc. That is, it may be a natural outgrowth of the mechanism he proposes, rather than a "fleeing" as such. The resolution of this is an empirical question, however.

Alonzo's theory also aids in interpreting the results of Schnore,[2] whose analysis of 1960 census data for 200 urbanized areas revealed that sheer age of settlement was the best predictor of the direction of city-suburban differentials in socio-economic status. Complementary to this, Berry[3] states

---

4 R.F. Muth, "The Spatial Structure of the Housing Market," RSA_P&P: 7 (1961), pp. 207-222.

1 Alonzo, op. cit., pp. 140-142.

2 L.F. Schnore, "Socio-economic Status of Cities and Suburbs," ASR: 28 (1963), pp. 76-85.

3 Berry, op. cit., p. 121.

slope than those with higher income.[1] As contrasted with indifference curves generally, the lower the bid-rent curve, the greater the satisfaction (curves belonging to one household are parallel). When the actual rent structure, per standard unit of land, is superimposed upon the family of bid-rent curves, the point of tangency of the rent curve and one of the bid-rent curves determines the equilibrium position of the household. Since those with lower incomes have steeper curves, they will tend to locate closer to the center than those with higher income and flatter curves. In addition, since the price of a standard unit of land declines with increasing distance from the regional center the amount of land which can be purchased with a fixed proportion of income increases, and residentail density would be expected to bear an inverse relation to distance.

There is general support for Alonzo's basic model. The conceptions of the earlier rent theorists, from which this work is derived, concerned the dominance of centers of various sorts.[2] Alonzo himself has tested it with the regression of expenditure for land upon income and distance from the core with Philadelphia data.[3] While the regression was for aggre-

---

1 Vernon has observed, however, that those whose income level gives them considerable freedom of choice have chosen every combination of access and spaciousness. This should warn against too simple an interpretation of preferences. See R. Vernon, Metropolis, 1985, Doubleday (Garden City:1963),p. 186.

2 W. Alonzo, Location and Land Use, Harvard U.P. (Cambridge:1964), Chapt. 1, provides a concise summary of this and other material.

3 Ibid., Appendix C.

gate census tract data and the sample small (15 tracts), the results supported his conception, with income having a positive coeffieient and distance a negative one. Brigham[1] and Frieden[2] have investigated land values specifically, and both have found a rent gradient which declines from the center of the region. However, Brigham, in a regression analysis, investigated this gradient by sector as well, and found that the regression of land values upon distance was not stable from sector to sector.

Urban geographers have documented population density as a function of distance from the city center, and have shown that a negative exponential formulation fits density patterns in a variety of contexts such as size of city and region of the world. In fact, Berry asserts: "No city has yet been studied for which a statistically significant fit of the expression $d_x = d_o e^{-bx}$ does not obtain."[3] Muth has pointed out that great variations in the fit occur for 46 American cities in 1950. Lower transport costs, larger population, higher average income, greater suburbanization of manufacturing, and greater proportions of substandard units in the central city tend to depress the gradient.[4] This is not surprising, in terms of Alonzo's theory, as central city deter-

---

1 E.F. Brigham, "The Determinants of Residential Land Values," Land Economics: 41 (1965), pp. 325-334.

2 B. J. Frieden, "Locational Preferences in the Urban Housing Market," JAIP: 27 (1961), pp. 316-324.

3 B.J.L. Berry, "Cities as Systems within Systems of Cities," Ch. 6 of J. Friedmann & W. Alonzo, eds., Regional Development and Planning, MIT (Cambridge:1964), p. 120. Here d is density, x is distance from center, d_o is density at center, and b is the density gradient.

that the gradient of density is a function of city size, and central density is a function of the city's age. Older age, in part, determines the type of structure built at the center as well as the possibility of mass transit facilities[1] and both would tend to increase density at the center. As the gradient constant and central density are the two other variables in the density-distance function cited above, this would tend to shed further light on Alonzo's formulation, as well as to fit Schnore's findings into Alonzo's framework.

Kain's[2] formulation of residential location is similar to Alonzo's, with the primary difference of "multiple centers" of a sort. For here, the workplace is the point of reference for each locatee, with rent paid for space of stated quality decreasing monotonically from the workplace, while transport costs increase monotonically. There is other work which supports such a conception. Carroll[3] found a tendency for workers to minimize distance to work. Lowry's work on population parameters for the Pittsburgh model also supports this, and bears on our later discussion as well.[4] His "trip distribution index" standardized the proportion of workers travelling from residence tract to work tract by the proportion of

---

[1] Schnore found that age of over 200 American metropolitan areas was the best predictor of mass transit usage. See "The Use of Public Transit in Urban Areas," in L.F. Schnore, _The Urban Scene_, Free Press (New York:1965), pp. 311-323.

[2] J.F. Kain, "The Journey-to-Work as a Determinant of Residential Choice," RSA-P&P:9 (1962), pp. 137-160.

[3] J.D. Carroll, Jr., "Some Aspects of the Home-Work Relationship of Industrial Workers," _Land Econ._:25 (1949), pp. 414-422.

[4] I.S. Lowry, "Location Parameters in the Pittsburgh Model," RSA-P&P:11 (1963), pp. 145-165.

workers travelling from residence tract to work tract by the proportion of employment in the work tract.[1] Distance travelled to work could be fitted to this index with an inverted power function for total work trips and work trips by occupational grouping. As the index was calculated with each of thirteen residential areas as a "center" for its residential population with respect to job locations, these findings give support to the importance of work-place in determining location. In addition, when the workers were stratified into four broad occupational groupings (professional-manager, clerk-sales, foreman-operatives, labor-service) the exponents of the functions ranked in the same order as the occupational groups' social status; that is, the lower the status of the group the closer the mean distance to work.

Duncan[2] examined work-residence separation for Chicago (1951), and her results also support Kain's formulation. Her tabulations of mean distance to work by occupational grouping (male and female, white and non-white) reveal an overall ranking which corresponds to the ranking of occupations by social status, with the exception that service workers are somewhat more removed from their workplaces than operatives. There is also a direct relation between income and distance to work, except for the highest income class

---

1 If W is resident population, E is employment population, and i & j are tract of residence and employment respectively,

$$T^*_{ij} = \frac{W_{ij}}{\sum W_{ij}} \Big/ \frac{E_j}{\sum_j E_j} \quad \text{and } T = 100 * \frac{T^*}{\sum T^*} .$$

2 B. Duncan, "Factors in Work-Residence Separation...", ASR: 21 (1956), pp. 48-56.

($100+/week). This ranking also holds for whites. For non-whites two lower income classes ($65-74,$55-64) are furthest separated from their place of work; additionally, except for the two highest and the lowest income classes, the non-whites' average distance to work is from ½ mile to over 1½ miles longer than whites. Further, non-white blue-collar workers are from ½ mile to 2 miles further removed from their work-places than are whites. These groups are probably the "reverse commuters" of which both Hoover and Vernon[1] and Meyer, Kain and Wohl[2] speak. That is, they are locked into central-city ghettoes, while their jobs are dispersed in the region. The effect of the segregation of a group is graphically seen in these results, and is relevant to our discussion below.

Reeder,[3] in a study of a much smaller area, Spokane, reports that no significant differences were found in a comparison of income with the length of time or the cost of the work trip; a comparison of occupational grouping with the length of time of the trip was significant, but the relation was inverse. These results are somewhat contrary to expectations, particularly with regard to income, but may relate to the smaller size of the region studied, the estimates obtained, or the small sample size.

---

1 E.M. Hoover & R. Vernon, Anatomy of a Metropolis, Doubleday (Garden City:1962), p. 168.

2 J.R. Meyer, J.F.Kain, M. Wohl, The Urban Transportation Problem, Harvard U.P. (Cambridge:1965), pp. 146-155.

3 L.G. Reeder, "Social Differentials in Mode of Travel, Time and Cost in the Journey to Work," ASR:21 (1956), pp. 56-63.

Kain goes on to present some quite interesting empirical material for Detroit concerning occupation, family-size, and space-consumption differences. The results generally tend to support his hypothesis of the trade-off between transport costs and site rents. One anomaly which occurred when he looked at residence by occupation, with workplace controlled, was that a "...lower percentage of sales workers and clerical workers reside in inner rings than their incomes would suggest..." and a "...higher proportion of operatives and craftsmen reside in inner rings."[1] One possible explanation for these discrepancies, which Kain proposes, is in terms of a greater female employment rate for the sales and clerical workers grouping (the tabulations were for total employed). That is, if we assume that female employment is often of a supplementary nature, possibly taken after the location of the household, the residential distribution of women might be reflecting, to a degree, the location of work of the husband. While this is plausible, it would also seem that, for secondary wage-earners, the place of work might be conditioned by the place of residence. It would also have to be demonstrated that secondary wage earners are concentrated more heavily in those households where the head already has a high income. Another possible explanation may be due to the segregation of non-whites in the central city. Elsewhere,[2] it is pointed out

[1] Kain, op. cit., pp.148-149.

[2] Meyer, Kain, and Wohl, op. cit., p. 145.

that centrally-employed whites may, in effect, have to commute farther than they "ordinarily" would if non-whites are centrally segregated, as well as suburban-employed non-whites having to commute farther than they "ordinarily" would, because of the imperfection in the market. Since the Detroit data were not controlled by race on this comparison, this explanation can be questioned as not totally adequate.

Another factor, which we will pursue, is that clustering by social class (white-collar, blue-collar) is at work to pull lower white-collar workers outward. It has been shown, for example, that for Chicago in 1940 the rent-income ratio for clerical and sales workers was higher than that for foremen-craftsmen.[1]

Several studies have demonstrated clustering, as such. Feldman and Tilly[2] investigated clustering of occupational groups in Hartford, Conn. (1950) census tracts by computing Kendall's tau for each group with every other group. The line between positive and negative correlations for occupational groups ordered by "prestige" is seen to be between clerical and craftsmen categories (the correlation between the two groups is near zero at .06). That is, all groups below the line (prestige-wise) correlated negatively with all the groups above the line. Eighteen of the 28 correlations

---

1 O.D. Duncan & B. Duncan, "Residential Distribution and Occupational Stratification," AJS:60 (1955), pp. 493-503.

2 A.S. Feldman and C. Tilly, "The Interaction of Social and Physical Space," ASR:25 (1960), pp. 877-884.

were above .30, twelve of these were above .50. Median school years and income were also correlated with occupational group. The rank correlation of median school years with the occupational categories was found to be of greater absolute value than median income for all but the service category. The rank intercorrelation of income and education was .54. They independently partialled the effect of income, then of education, and found that education accounted for more variance in the occupation-by-occupation correlations than income, i.e., the partials were lower when education was removed.

It would appear from the results of this study that a "social class" effect, represented most strongly by a white-collar/blue-collar dichotomy, is at work in the determination of residential patterning, and that this effect is not simply a result of differential income, but concerns attitudes toward type of neighborhood and neighbors desired, as well as a host of other effects upon family size, lifetime income cycles, etc.[1] From such a point of view, it would be expected that low-status, white-collar workers, while lower in income than high-status, blue-collar workers, would have a residential distribution more similar to other white-collar workers than to blue-collar workers of similar income. The converse of this would be expected to hold true for high-status, blue-collar workers as well.

The study by the Duncans[2] also throws light on occupa-

---

1 J.M. Beshers, Population Processes in Social Systems, Free Press (New York:1966), has a good discussion of these effects.

2 Duncan & Duncan, op. cit.

tional clustering. They considered the occupations of males, ages 14+, for 1950 Chicago data. Four indices were computed -- dissimilarity, segregation, low-rent concentration, and centralization -- which represent correlation coefficients of sorts. The index of dissimilarity is an "inverse correlation" of the distribution by tract for two occupational groups. The results in general showed a white-collar/blue-collar split, but, as contrasted with the Feldman-Tilly findings, clerical workers were more dissimilar to other white-collar categories than to blue-collar categories. Their "index of low-rent concentration," however, had magnitudes which were ordered according to the accepted prestige ranking of the occupations, with the exception that service workers seemed to fall between the clerical and skilled categories. Their index of centralization gave results ordered generally the same as the prestige ranking, with the exception that clerks and foremen were reversed -- that is, foremen were less centralized. As these tabulations were for males, Kain's explanation for the reversal found in his data would seem to be supported. The important qualification to the Duncans' results is that place of work could not be controlled, as for the Detroit study. If, as is reasonable,[1] foremen lived and worked on the periphery and near the core, while a high proportion of clerical workers lived and worked near the core, the Duncans' results and Kain's results might

---

[1] L.K. Loewenstein, "Industry of Employment...", AJE:24 (1965), pp. 157-162, has results supporting such a conception, as he found manufacturing workers heavily distributed in the outer rings of 18 American cities.

not be incompatible.

Finally, the Duncans found a U-shaped pattern for their index of residential segregation. Laborers were most highly segregated, followed by professional, manager, sales, and service categories (of about equal segregation), followed by semi-skilled, skilled, and clerical categories. The ranking in order of prestige -- professional, manager, sales, clerical, skilled, semi-skilled, service, laborer -- produces a trough-shaped curve, with clerical workers at the bottom.[1] The Duncans note that the amount of residential segregation is greater for those groups with clearly-defined status.

The possible consequences of this are worth exploring. It has been found that distinctions as to the position of occupational groupings on an hypothesized uni-dimensional scale of prestige are better made by those at the ends of the scale, than those in the middle.[2] This pattern bears an interesting resemblance to one hypothesized attitude mechanism -- that of assimilation-contrast -- concerning a point of abrupt discontinuity connected with judging likeness or difference. That is, up to a point, dissimilar objects will seem to be like the reference object (in this case, persons' statuses are the

---

1 A replication of the Duncan study by Vyeki with Cleveland data for 1950 and 1960 yielded results similar in pattern and magnitude to that of Chicago. In addition, the patterns persisted over the decade, and labor became more distinct. See Vyeki, E.S., "Residential Distribution and Stratification, 1950-60", AJS:69 (1964), pp. 491-498.

2 J. Beshers, Urban Social Structure, Free Press (New York: 1962), pp. 140-141 has a discussion of this.

objects), and, after this point, abruptly, different. Perhaps, those in the middle of a structure are not sufficiently far from anything to make sharp contrasts, while those at either end are.

Such a mechanism might be at work, in part, in the perception of social classes by those at various points in the structure. Laumann, in his discussion of these questions[1], provides an explanation in terms of an hypothesis well-known in voting studies -- the cross pressure hypothesis[2] -- which has implications for residential segregation. Laumann found that those men with high "associational congruence"[3] also tended to be more class conscious, as well as to be more likely to express a political party preference. These results indicate the possibility that persons in a homogeneous environment reflect it to a greater degree than those subject to competing points of view and, possibly, divided commitments. The question of whether association determines attitudes or attitudes determine association is not settled here, or elsewhere, although research on group dynamics would indicate that

---

1 E.O. Laumann, Prestige and Association in an Urban Community, Bobbs-Merrill (New York:1966), Chpt. 8, passim.

2 The body of literature on cross-pressure is fairly large, and, while it is certainly not a completely adequate explanation, a good bit of support for it can be found. Ibid., p. 135, footnote 9, lists several of the standard references.

3 Ibid., p. 132. "Associational convergence" measures the amount of association with those of similar occupational background as oneself, by Laumann's definition.

the effects of association upon attitudes is not negligible.[1]

Beshers has argued that residential location and social structure tend to complement and reinforce each other -- social structure influences residential location, which in turn influences and maintains social structure.[2] Aside from the study of marriage patterns, this question has received little systematic attention for metropolitan sub-areas. There is an abundent literature on small towns, including studies of interaction patterns among members of different occupational groupings, however, their applicability to the urban setting may be questioned. The incidence of non-voting or Independent voting might be employed as an indirect index of the integration of the tract in rough empirical work. However, while voting seems to relate to the cohesion of the sub-area[3] the voting statistic is "contaminated" (for this purpose) by a generally higher incidence of non-voting among blue-collar groups, and would give little information concerning the amount of contact among status unequals.

Several studies have indicated that contact with neighbor, while less than contact with co-workers, friends generally, or relatives, is still considerable, and does increase with the

---

[1] Here, the body of literature is immense. Homans discusses a large number of studies along this line. G. Homans, Social Behavior, Harcourt, Brace, & World (New York: 1961).

[2] Beshers, op. cit., Chapt. 6, passim.

[3] S. Greer, "Citizen Participation and Attitudes", in Bollens, J.C., ed., Exploring the Metropolitan Community, U. of Calif. Press (Berkeley & Los Angeles: 1961).

"familism" (single family dwellings, large proportion of children, etc.) of the area.[1] In Greer's study, for example, neighboring was lower in areas characterized by a higher proportion of multi-family dwellings, lower proportion of children, and the like, but was not insignificant, even here. This indicates that informal neighborhood association is not wholly lacking in most parts of the city, but says little about the character of this association. Association among status unequals may be more prevalent in areas of high familism, for example. As results relevant to such a question were not presented in the articles cited, nothing can be said on this point.

Laumann's study is the only source of information with direct relevance. Laumann did not find a great deal of close association between neighbors of highly incongruent status[1] in his sample of Cambridge and Belmont, Mass., residents, although he gathered data on only the immediate neighbors of the respondents. Causal chatting only or no association was found to relate to incongruent status of neighbors, while a "good friend" relationship was found more often among neighbors of like status. These results were not further controlled by the occupational group of the respondent, so it is impossible to say whether incongruent associations with neighbors were more frequent for, say, clerical workers than for professionals.

---

1 Bell, W. and M.D. Boat, "Urban Neighborhoods and Informal Social Relations", AJS:62 (1957), pp. 391-398.
Greer, S. "Urbanism Reconsidered: A Comparative Study of Local Areas in the Metropolis", ASR:21 (1956), pp. 19-25.

2 Greer, op. cit., p. 134.

Since clerical and skilled workers had the lowest proportions of neighbors in their own occupational category,[1] they could have been those biasing the results toward no association with incongruent neighbors, or they could have been the strongest contributors to association -- we have no way of telling.

Due to this lack of research results on association within an extended area (Census tract size, say) it is still an open question how residential location interacts with contact among status groupings. In view of this, it is perhaps worth noting that the more uniform distribution of clerical workers, for example, may be the result of their, in part, filling in both economic and social crevices. That is, after the rent theorists, those with both low and high resources may exert strong locational preferences -- low income workers needing to live close to work, and thus occupying small amounts of high-priced land, and those of high income desiring other benefits, and paying for them with higher commuting costs. Those in the middle may take the residual. Again, after the social class theorists, it may be that location decisions are shaped by class identification, which seems to be high at either end of the scale, while those in the middle, with lower identification, do not seek homogeneous neighborhoods. On the other hand, it is not unreasonable that, given such a patterning, the nature of possible contacts in each neighborhood will tend to reinforce class conceptions, if the neighborhood is homo-

---

[1] *Ibid.*, p. 71.

geneous, or moderate them, if the neighborhood is heterogeneous. This would be so if we can assume that interaction, itself, and among status unequals, does take place in a heterogeneous neighborhood.

Granting that a rent gradient exists, and the strong effects of income and workplace, it would yet seem that clustering occurs for social class reasons. The next question would be, how will a "social class bias" manifest itself? Aside from the possible outward shifting of lower-income white-collar workers, we would suggest that this effect will come out in a sectorial pattern. Several studies are suggestive of this. Hoover and Vernon, for example, found that "...professional and managerial people, though having rather different patterns of work-place concentration, choose very nearly the same list of counties for residential concentration."[1]

Kish[2] in a study of eleven large metropolitan areas and two small states found differentiation among communities within suburban rings on a variety of census variables and voting. His measure was the intraclass correlation coefficient, which, in this context, is the ratio of the variance between communities in a ring to the total variance in the ring, adjusted for the variance within communities. High positive values thus indicate that the variance is a result of the "specialization," in a sense, of communities, while low or negative

---

[1] Hoover & Vernon, op. cit., p. 151. The emphasis is theirs.

[2] L. Kish, "Differentiation in Metropolitan Areas," ASR:19 (1954), pp. 388-398.

values indicate a random sorting of individuals into communities. His result most useful here is that communities within rings do seem to be specialized. There is a gradient pattern to this differentiation, as well, since the magnitude of the coefficient decreased outwards from the center of the regions. The gradient was not smooth, but fell off sharply, after a point, then resumed a slower decline. The band of higher differentiation seemed to be wider for larger areas than small ones. These results parallel those concerning population density, discussed above.

In a significant article, Anderson and Egeland[1] examined, by the use of social area analysis and the analysis of variance, both the radial and concentric patterning of attributes of the population. Before discussing their results, a few brief notes should be included to clarify the nature of social area analysis.[2] Three indices are involved -- social rank, urbanism-familism, and segregation -- which are generally computed at the census tract level by the scoring of census tracts as a whole. The social rank construct is simply a combination of occupation and education scores, the urbanization construct is a combination of the fertility, women-in-the-labor-force, and single-family-detached-dwelling ratios. The segregation

---

1 T.R. Anderson and J.A. Egeland, "Spatial Aspects of Social Area Analysis", ASR:26 (1961), pp. 392-398.

2 See the abstract of their book in E. Shevky and W. Bell, "Social Area Analysis," pp. 226-235 in G.A. Theodorson, ed., Studies in Human Ecology, Harper & Row (New York:1961).

index is designed as a measure of ethnicity (of "new" immigrant groups and non-whites).

Several factor analytic studies of census-tract data have tended to support the use of these indices. Anderson and Bean[1] found the general schema verified for Toledo, although urbanism and familism emerged as different factors. Van Arsdol, and others,[2] in a study of ten American cities, found the factor loadings generally supported the conception. Again, however, discrepancies arose in several cities; fertility, in two cities, and single family dwelling units in another, were associated with social rank more strongly than with urbanism/familism. This indicates that more fundamental relations exist, and that the Shevky schema captures only their most common manifestations.

Beshers,[3] working with a larger set of variables for Cleveland, found four factors strong enough to support interpretation. The first two and fourth were identifiable as social status, young family, and race, while the third represented high proportions of females and residential stability.

---

1 T.R. Anderson & L.L. Bean, "Shevky-Bell Social Areas...", SF: 40 (1961), pp. 119-124.

2 M.D. Van Arsdol, S. Camerilleri, C.F. Schmidt, "An Application of the Shevky Social Area Indexes to a Model of Urban Society," SF: 32 (1958), pp. 26-32.

3 J.M. Beshers, "The Construction of 'Social Area' Indices: An Evaluation of Procedures," Paper read at the Am. Stat. Assoc., Social Statistics Section, December, 1959.

Again, Shevky's urbanism/familism index is found to be suspect. In addition, Beshers' analysis of the original correlation matrix revealed important linkages between the general groups of variables -- an aspect which is overlooked if one simply considers orthogonal factors and their loadings. That a correlation matrix can be represented by (perhaps a smaller number of) independent components is, to be sure, an interesting result, but it is not at all clear how seriously this result should be taken. To calcify it into "reality" of some sort would seem to be unjustified.[1] This is not to deny the utility of factor analysis in data reduction, in providing provisional hypotheses, or in the formation of homogeneous groups of observations (either from an inspection of the factor scores or from factoring a transposed -- variable by observation -- data matrix), but to indicate that the results of a factor analysis, more than any other statistical technique, must be confirmed in other ways.

Clarke,[2] in a principal components analysis of Boston data (by town), found four components. The First two, urbanism/suburbanism and income/occupation, correspond to Shevky's urbanism/familism and social rank. Two others appeared, which Clarke called statility/industry I & II, which represented "lower-white-collar stability and instability," would seem to be related to Beshers' "female-stability." Again, it would appear that the Shevky typology omits important variation, particularly in terms of the circulation of house-

---

[1] See Section 4 of H. Hotelling, "The Relations of the Newer Multivariate Statistical Methods to Factor Analysis," Brit. Journal of Stat. Psych.: 10 (1957), 69-79, for an argument against determining "genetic" factors from factor analysis.

holds. Other studies which we will not treat here, have been done. This sampling indicates, however, that the schema does have utility, subject to the qualifications above, for the analysis of urban areas and that, if it does not represent all of the necessary components it was designed to measure, it succeeds reasonably well.

Anderson and Egeland[1] examined four cities -- Akron and Dayton, Chio, Indianapolis, Ind., and Syracuse, N.Y. -- of between 200,000 and 500,000 population. Four 30 sectors were selected for each town, and four census tracts were selected, on the basis of distance from the city center, from each sector. An analysis of variance revealed that distance from the center (and inter-city variation, but not sector variation) was the important contributor to the variance of the main effects in the urbanization index. The social rank index (occupation and education), had its variation located in the sectors, while distance was significantly related to social rank for only one city.[2] Thus, by this test, social rank and urbanization appear as independent, additive components to the form of the city.

This ties in with the previous discussion in several ways. First of all, it tends to support Alonzo's conception of the trade-off between distance from the core and quantity

---

[2] W.L. Clarke, "Intra-Metropolitan Migrants and Town Characteristics," unpub. MCP thesis, MIT, 1967.

[1] Anderson & Egeland, op. cit.

[2] Since there were no significant first-order effects for urbanization, the three-way design could be tested. But, since city-by-sector interaction was significant for social rank, a two-way ANOVA was done for each city, to test social rank.

of land purchased, in that the social rank index consists, in part, of the single family dwelling variable. In addition, it demonstrates that, even with this source of variation taken into account, occupational differentiation takes place in a sectorial fashion. It should be pointed out that such variation could arise from different degrees of accessibility for various sectors, or from a sectorial patterning of industry.

The differential accessibility, particularly to the core, which sectors might have could result in a significant sectorial effect (within an analysis of variance) by pulling lower-income groups toward the center more in some sectors than in others. Other things being equal, this would result in spurious sectorial patterning, for, even if the proportion of an occupational group in each sector were a constant, and job location was similarly standardized, differential accessibility would still cause varying proportions of occupational groupings within any given ring. Since the areas were of medium size, accessibility may not have been an important issue, but this is open to question.

The sectorial patterning of industry would have the effect of pulling various occupational groups into a sectorial pattern, as well, and probably represents the more potent force of the two. Industry which is transport-sensitive, particularly to rail or water transport, will, by definition, tend to be strung along the network[1] which will probably depart radially from

---

[1] E.M. Hoover, The Location of Economic Activity, McGraw-Hill (New York:1948), p. 130.

the core. Loewenstein[1] has mapped and discussed the locational patterns for 44 small and medium-sized American metropolitan regions. His data reveal a strong sectorial pattern for this industry. In fact, of the eight industry categories he studied, only public administration did not have a sectorial patterning (or decentralized location). Even retail, where it was decentralized, tended to be patterned in a sectorial fashion. Lower-income workers would tend to locate more closely to their place of employment, and would thus be pulled into a sectorial arrangement and higher-income workers, while locating farther from work, would tend to choose areas not inhabited by industry of the noisome kind, such as traditional manufacturing concerns, and because the relative cost of land would be less.

In view of the sectorial patterning of work-places, it might seem as if much of the patterning by social rank could be accounted for on this basis. However, this is essentially an empirical problem, as no study has treated it adequately. In view of the obvious clustering of occupational groups at the census tract level, and the research results reported above, a sectorial class bias may be postulated. In part, this is likely due to patterning of workplace by sector. However, the hypothesis that households are simply generally constrained by the overall rent gradient of a region, roughly patterned by annular rings from the center, and by transport costs, acting to further determine the ring and the general

---

[1] L.K. Loewenstein, The Location of Residences and Work Places in Urban Areas, The Scarecrow Press (New York:1965), Chpt. 2, passim.

sectorial patterning, while social class considerations enter to determine the particular sub-area of residence, would seem to hold promise for the resolution of these various findings.

In summary, then, we have reviewed a variety of theories and findings concerning residential location in order to put the following empirical work into some perspective. Alonzo's theory of land rent, with the core of a region as the focus for a rent gradient, seems to provide an adequate basis for characterizing the overall structure of a region. Kain's formulation in terms of distance from workplace both complements and extends this to multiple centers. In view of the variation in the patterning of workplaces, this adds a sectorial patterning to the structure of a region. However, other evidence has indicated that these two forces alone may not be adequate in the explanation of patterning, particularly at the small-area level. In particular, social class considerations seem to operate to both partially disrupt "normal" market forces and to determine a pattern of clustering which follows a prestige ranking more closely than an income ranking. The implications of an interaction between market forces and class conceptions were treated briefly.

Household Migration

One question which must be posed in relation to intra-metropolitan migration is what has been variously termed the "stayer-mover" or "population-at-risk" problem. It may be posed simply - can the population be divided into two (or more) categories with different, but internally consistent, probabilities of moving. If we can identify stayers (p=0) and movers (p>0), then turnover figures are put in a new, and proper, perspective, and the process of prediction, as well as analysis, takes on new life. Two sets of relatively independent processes are then involved: one, those processes tending to pry the individual or household loose from the immediate environment and raising the chance of exposure to the other, the allocation process. While several models of the allocation process only have been proposed - those of Stewart/Zipf and Stouffer, for example - at least two model-making attempts have faced this question squarely. Blumen, Kogan, and McCarthy[1] attempted to predict labor turnover with a simple Markov process but found that much closer estimates of turnover (that is, less turnover) were obtained when the transition matrix reflected a stayer and a mover group. Porter[2] has also proposed a model of migration which treats the flow between two points as probablistically determined by job vacancies at

---

[1] I. Blumen, M. Kogan, and P.J. McCarthy, The Industrial Mobility of Labor as a Probability Process, Cornell U. Press (Ithaca:1955).

[2] R. Porter, "Approach to Migration Through its Mechanism," Geografiska Annula:38 (1956), pp. 317-343.

destination, applicants at source, and, in a sense, the "intervening opportunity" of a vacancy being filled by an applicant from between the two places.

At the inter-regional level, Lowry[1] has demonstrated that economic conditions at the destination are highly correlated with size of flow, whereas those at the source are not. A similar effect was also noted by Gallaway.[2] That some demographic classes have high rates of inter-metropolitan migration while others do not is commonly admitted.[3] That a link between inter- and intra-area mobility could be established is evidenced in data gathered by Goldstein.[4] For three time periods, by tracing individuals' residences over time through city directories of Norristown, Pa. he was able to establish that not only was the out-migration rate higher for recent immigrants (and vice versa) but that those who entered and then left the area in a period also had a larger mean number of residences (3-4) than any other group.[5] Incidentially, it

---

1 I.S. Lowry, Migration and Metropolitan Growth, Chandler (1966), Chapter 2.

2 L.E. Gallaway, "Labor Mobility, Resource Allocation, and Structural Unemployment," AER:53 (1963), pp. 694-716.

3 See Lowry, op. cit.; J.M. Beshers and E.N. Nishiura, "A Theory of Internal Migration Differentials," Soc. Forces:39 (1961), pp. 214-218; J.D. Tarver, "Occupational Migration Differntials," Soc. Forces: 43 (1964), pp. 231-241.

4 S. Goldstein, "Repeated Migration as a Factor in High Mobility Rates," ASR:19 (1954), pp. 536-541, and Patterns of Mobility, 1910-1950, U. of Penna. Press (Philadelphia: 1958), pp. 210-214.

5 Patterns... , pp. 210-214.

was found that those residents who out-migrated and those immigrants who stayed had an equal mean number of residences (near 2) and that this was larger than those in continuous residence (1.3-1.5).

Very little research has been done concerning the measurement of "intra-regional migration potential", although guidelines for analysis have been presented by Beshers.[1] Briefly, he considers job-related constraints - the linkage of jobs to areas as constraining certain occupational groups (such as small entrepreneurs), the career as determining the possible style of life and expected income patterns and income as putting an upper bound upon the type of housing obtainable. Household-related constraints include the number and spacing of children, need for schools and recreation, and status seeking propensities, particularly for those families with daughters.

One study has related the journey to work and migration,[2] and supports one part of the job-related constraints theory. It was found that the greater the distance moved, the less tendency there was for the worker to retain his job at the previous place of residence. Studies by Rossi[3] and Ross[4] have focused on life cycle and attitudinal factors, particularly family-cycle variables. These studies support parts of the household-related constraints theory.

In conclusion, it would seem to us that family and career cycles are the primary driving forces behind residential shifts. Other factors, such as urban renewal, racial pressure, and tax rate shifts, to mention a few, while important, would seem to be dependent upon the life-cycle factors. Unless a family is ready to move, and financially able to move, strong coercion would be required to move them. While past studies have shown a connection between the origin and destination of movement in terms of social characteristics,[5] the role of "push" and "pull" factors has not been thoroughly studied.[6] To determine if there is indeed a push from certain types of communities which is more than a simple "prying loose" do to life cycle factors, it would have to be demonstrated that household types which are generally low in terms of migration rates are, in fact, leaving these areas at a greater rate than would be expected. In either case, the allocation mechanism would yet be expected to operate in terms of the residential location factors discussed above.

---

[1] Beshers, Population Processes..., op. cit., chapter 5.

[2] S. Goldstein & K. Mayer, "Migration and the Journey to Work," SF: 42 (1964), 472-498.

[3] P. Rossi, Why Families Move, Free Press (New York:1955).

[4] H.L. Ross, "Reasons for Moves to and from a Central City Area," SF:40 (1962), 261-263.

[5] E.C. Rust,"Intra-Metropolitan Migration: Six Boston Area Municipalities," Unpublished M.C.P. thesis, M.I.T. (1963)
G.P. Leyland, "Migration to and within a Small Area," Unpublished M.C.P. thesis, M.I.T. (1966).

[6] Although Henny has investigated the pull factors for a selected category of household. See L.M. Henny, "Differential Educational Opportunities and Residential Choices..." Unpublished M.C.P. thesis, M.I.T. (1966).

Chapter III

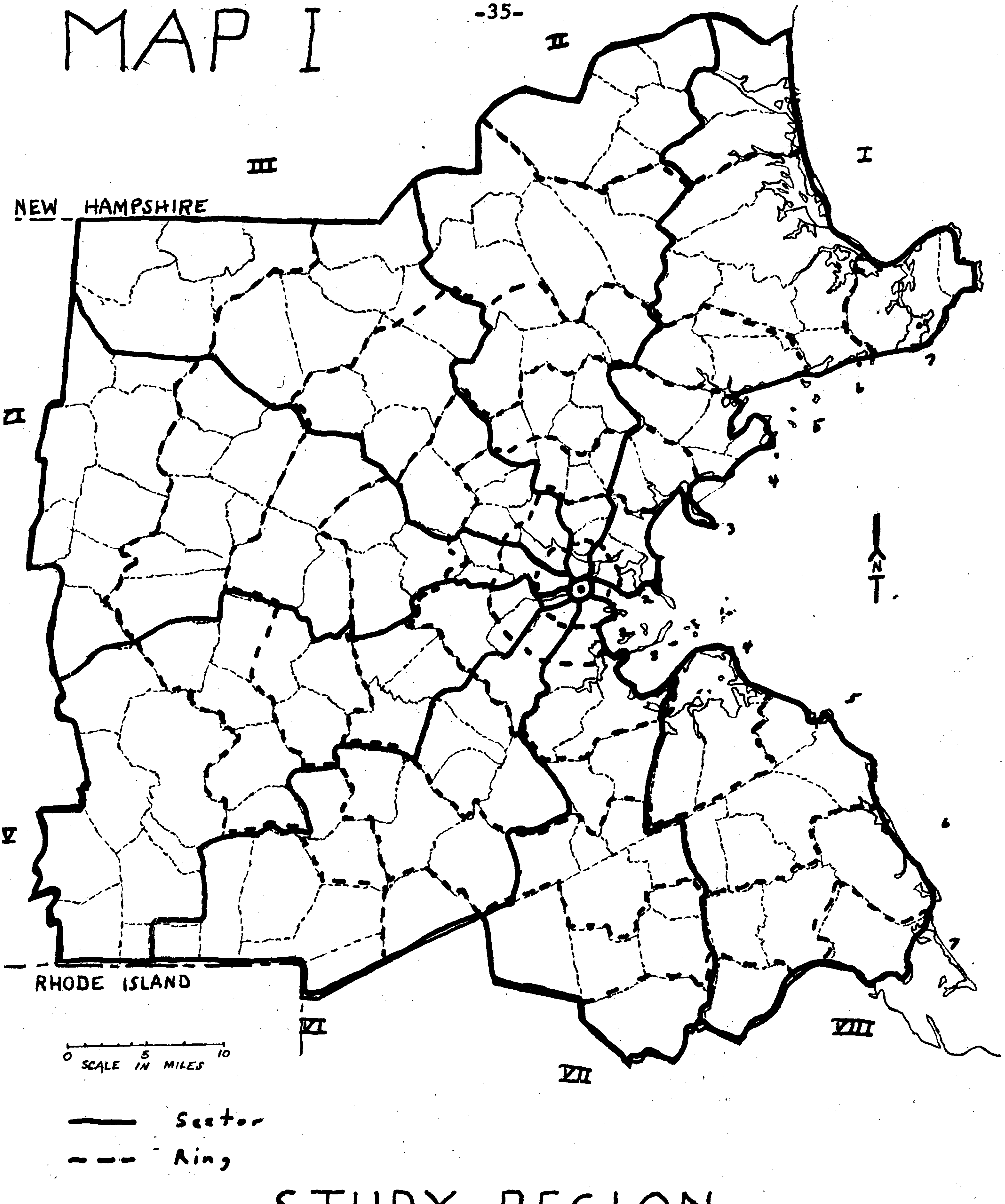
FACTORS INVOLVED IN LOCATION AND MIGRATION: EMPIRICAL WORK

Canonical Correlations

Method for Canonical Correlations:

The data for this part of the analysis came from a tape containing population counts (and medians) for a variety of variables at the census tract level. The study area was that shown in MapI - an area with roughly a 35 mile radius consisting of 152 towns and 598 tracts. The region was subdivided into a pattern of sectors and rings to allow comparison of sub-area characteristics. Several criteria were laid out for the drawing of these boundaries. First, the classification was to be done on an areal basis, without reference to population distribution. If a tract or town was very nearly bisected by an "absolute" ring or radial, an arbitrary decision was made as to which area would include it. No attempt was made to classify homogeneous areas by statistical criteria, although this would be an instructive exercise. Second, since certain of the data contemplated for use was only available at the town level, and, at the outside of the region, the tracts consisted of whole towns, town boundaries were respected wherever possible. Third, as the sectorial part of the geographic analysis was exploratory, the sectors were to be of roughly the same width. Several widths were tried for the sectors; 35-40 degree and 50 degree widths were found to fit the town boundaries "best" while fitting into the 300 degrees comprising the region. It was felt that eight sectors, rather than six, would not be unwieldly, and would capture migration streams more accurately, while also representing "natural" areas, such as the North Shore, in

MAP I



STUDY REGION

a better fashion. Finally, the widths of the rings were set to roughly correspond with those used in other studies, so as to allow comparability.

The boundaries of the rings and sectors are displayed in Map I. The rings are of radii 1, 3, 6, 10, 15, 20, and 27 miles, with the outside ring comprising the rest of the region, while the sectors are from 35-40 degrees wide. Sector 0 and Ring 0 are the same, and lie within one mile of the center of the region.

Tracts were omitted from the analysis entirely if they had fewer than 50 persons, an institutional population of more than 25%, or missing data in any of the variables. In all, thirty-one tracts were omitted for one or more of these reasons. Twenty of these tracts were within three miles of the center of the region, (Rings 0 and 1), while the others were scattered over the rest of the region; for this reason, Rings 0 and 1 were grouped together in the computations.

The primary objective was to investigate the clustering of occupational groups within tracts to determine the nature of this grouping in the region as a whole and in its parts, as well as attempting to assess, if crudely, how much of the variance could be accounted for by indices of purchasing power and "life style". The variables initially employed were median income of families and unrelated individuals, median education of persons aged 25+ and employed males in eight occupational groupings as a percent of the males, 15+ years, in the tract.

Other variables were added in later computations.

The use of medians, rather than means, was an unfortunate necessity, as the standard error of a median can be shown to be, for N greater than 100, about 1¼ that of the mean. In terms of sample size, this means that the sample must be one half again as large to achieve the same precision with a median as with a mean. No workable alternative was available, however, and, as the number of tracts, even in most sub-areas, was fairly large this should not be a serious bias. The use of income as an index of raw purchasing power is more defensible than the use of education as a proxy for a large complex of other variables, including life style preferences, family size, aspirations for children, ethnicity, etc. However, these two had the fewest missing observations and present the clearest contrast conceptually of any of the set of variables with which we had to work.

Since the emphasis was upon the location of residence, and since the employment of females is often of more incidental importance in the location decision of an "average" household (the argument being that their employment is often of a supplementary nature, and by hypothesis, after the fact of residential location[1]), only male employment figures were included in the percentages. The category of household services was dropped

---

[1] Professional-technical, manager-official-proprietor, clerical, sales, craftsman-foreman (skilled), operative (semi-skilled), service - except private household, and laborers (unskilled).

[1] See Kain, op. cit., pp. 146-147.

due to the low incidence of males (a mean of 7 males per tract).

The percentages were left in their raw form, although it has been suggested that a transformation of the form $q = \sin^{-1}\sqrt{p}$, where p is the proportion, tends to make the variances independent of the mean.[1] This is so since a very low or very high proportion can be determined with greater accuracy than one around 50%. The square root corrects for the dependence, and the angular transformation corrects for the symmetry about 50%. While this transformation step would be desirable, there were several reasons why it was not done. In the first place, the percentages of the occupations in a tract were generally below 20% - few tracts had percentages of any occupation above 30%. In addition, "rigorous" hypothesis testing was not our aim here; the untransformed percentages should not have too disturbing an effect on the correlations.

Product moment intercorrelations were computed for this set of ten variables, at the tract level, for the region as a whole, and for each sector and ring, with the exception of Rings 0 and 1 being grouped together, as mentioned previously, and Sectors 7 and 8 also being combined, due to the small number of cases in Sector 8. Canonical correlations were then run on these sets of intercorrelations, with the occupational variables on one side, and income and education on the other. Distinctions as to dependent and independent variables are not really possible with this technique as both sets are independ-

---

[1] Quenouille, M.H., Introductory Statistics, Butterworth (London:1950), pp. 168-170.

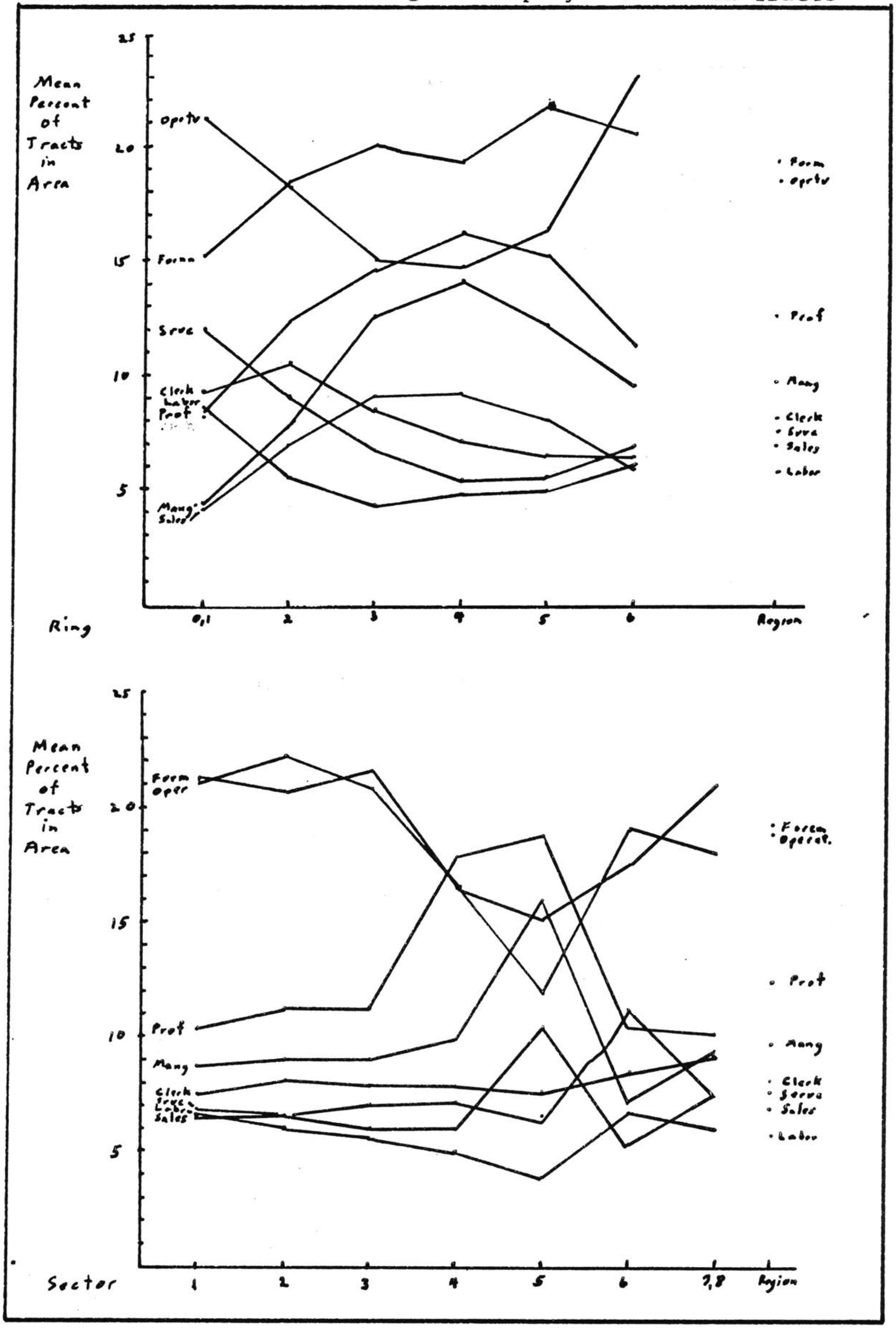
ent variables. When dealing with aggregate data of this sort, it is meaningful to ask whether a certain observed pattern of occupational distribution "causes" (in a weak sense, from my point of view) a certain distribution of income and education, as well as whether education and income "cause" (in a somewhat stronger sense) a distribution of occupations. As a by-product of the solution of the canonical correlations, the principal components of the occupation set were printed out; it is of some interest to look at these also, and they will be discussed briefly.

Distribution of Percentages:

As an initial step, the means of the percentages for occupational groups in the tracts were plotted for each occupation by ring and by sector. This was not done to provide an estimate of the actual percentage in each sub-area, since it would do so only if the tracts were of equal size, but only to compare the measure of central tendency of the statistic for each occupation. As the emphasis here is on clustering by tract, these means are more informative than aggregates for the sub-regions. The results are plotted in Graph I. A complementary pattern can be observed for the rings as well as for the sectors.

For the rings, "white-collar" occupations generally tend to slope away from a peak at Ring 4, while "blue-collar" occupations tend to slope upward from a trough in this same region. The two exceptions to this are clerical workers, whose mean

GRAPH I - Mean Percentage of Employed Males in Tracts



percentage tends to peak at Ring 2 and foremen, whose mean percentage tends to rise, dip at Ring 4, rise and then fall. While these results are not controlled by place of work, they generally bear out a formulation which posits an interaction between location rents and transport costs - blue-collar jobs tend to be concentrated in the center and toward the outside of the region: their holders, generally lower-paid (see Table I), would tend to locate near to them. White-collar jobs tend to be relatively more concentrated in the center, and their higher-paid holders would tend to locate farther from them, i.e. in a middle band of residence. It is interesting to note that clerical workers, having the lowest median income of the white-collar group tend to be found in relatively higher percentages toward the center of the region, while foremen, with the highest incomes of the blue-collar workers tend to locate more in the fashion of white-collar workers.

The pattern of mean percentages by sector sheds further light on the distribution of occupations. Here, with the exception of the clerical category, the pattern is again complementary with respect to a white-collar - blue-collar dichotomy. In the northern and southern sectors (Sectors 1-3, 6-8) the mean percentages for each group are generally close to that for the region as a whole, while in the western sectors (Sectors 4-5, and to a degree 6) the means take a large jump for white-collar and a dip for blue-collar categories.

These results are in accordance with those of Clarke[1],

---

[1] Clarke, _op. cit._ See his "Combined Factor 2, Income-Occupation" map.

TABLE I

Median Yearly Income for Occupational Groups
Boston SMSA:1960

Occupation - Males	Income
Managerial (Non-farm)	7420
Professional	6741
Foreman	5378
Sales	5345
Clerical	4598
Operative	4501
Service	3864
Labor (Except farm & mine)	3543

Occupation - Females	Income
Managerial (Non-farm)	4095
Professional	3737
Foreman	3285
Clerical	3083
Operative	2475
Labor (Except farm & mine)	2051
Service	1697
Sales	1543

SOURCE: U.S. Bureau of the Census, U.S. Census of Population:1960. Detailed Characteristics. Mass. Final Report PC(1)-23D. U.S. Government Printing Office (Washington,D.C.:1962), Table 124, "Earnings in 1959 of Persons in the Experienced Labor Force, by Occupation..." pp. 23-415 to 23-417.

which revealed high factor scores for his income-occupation factor concentrated in a western wedge and in the middle rings. Low factor scores were distributed at the outside of the region, as well.

Correlations and Components:

Product moment correlations among the eight occupation variables and the income and education variables were computed, by tract, for the region as a whole as well as for sectors and rings. The intercorrelation of percentages is open to some question if the denominators are the same, as, depending upon the relations of the "rel-variances", $v = s/\bar{x}$, the ratio of the standard deviation to the mean[1] for each numerator and the denominator of the percentages, the correlation can be raised or lowered from what it would be if the numerators were initially uncorrelated.[2] Meyer and Kuh also discuss the conditions under which the correlation of percentages will approximate the correlation between the numerators with the denominator partialled from it. However, our interest was in the correlation of the "amount of space" in a tract taken up by the occupational groups, not in removing the effect of the size of the tract population, as such.

The alternatives to this procedure should be discussed briefly. One alternative is to abandon the correlational frame-

---

1 In Hanson, Hurwitz, and Madow, Sample Survey Methods and Theory, vol. I, Wiley (New York:1953), pp. 24ff.

2 See Kuh and Meyer, "Correlation and Regression Estimates When the Data are Ratios," Econometrica:23 (1955), pp. 400-416.

work altogether, and proceed to construct an index of dissimilarity, as was done by the Duncans. Aside from the need for a set of correlation coefficients to use in the canonical correlations, another reason militates against the Duncans' method. That is, they compute the ratio of the number of persons in a tract in an occupational category over the total for that category in the region. Since the denominators of the percentages are different, the results of a parametric or non-parametric correlation performed upon them would be, in general, different. More to the point, since the Duncans' index is one of displacement, that is, it states that X% of the population in one category would have to be redistributed to bring the percent of each in the same tract into agreement, it gives no information about the direction of relation present.[1] A well-chosen, nearly-uniform distribution for one variable could give the same displacement score as a perfect inverse relation. The correlation coeficient, on the other hand, does take the direction of relation into account, but since it is based on an assumption of linearity, a U-shaped curve for two variables would show up as a near-zero correlation.

The correlation matrix for the region as a whole (Table II) is representative of the general pattern for the sub-areas, and is of some preliminary interest. Due to the sample size, a correlation of greater than .11 is significantly different from zero. With the exception of the clerical category, the

---

[1] J.M. Beshers, E.O. Laumann, & B.S. Bradshaw, "Ethnic Congregation, Segregation, Assimilation, and Stratification," SF: 43 (1964), pp. 482-489, present a good discussion of the indexing problem.

TABLE II

Occupational Residential Correlations
Boston Study Region: 1960
Census Tract as Unit

Occupation	Prof	Manag	Sales	Clerk	Skill	Semi-Skill	Serv	Labor
Professional	1.00	.68	.55	-.19	-.38	-.76	-.33	-.55
Managerial	.68	1.00	.76	-.32	-.25	-.68	-.49	-.52
Sales	.55	.76	1.00	-.12	-.21	-.64	-.40	-.54
Clerk	-.19	-.32	-.12	1.00	.05	.04	.25	.01
Skilled	-.38	-.25	-.21	.05	1.00	.32	-.20	.07
Semi-Skilled	-.76	-.68	-.64	.04	.32	1.00	.11	.46
Service	-.33	-.49	-.40	.25	-.20	.11	1.00	.29
Labor	-.55	-.52	-.54	.01	.07	.46	.29	1.00

SOURCE: Census Tract Tape

NOTE: The number of the male labor force in a census tract is the denominator of the percentages in the correlations. The occupation variables are correlations of the percentage of each category of the labor force population in the tract. The skilled category represents foreman-craftsman. The semi-skilled category represents operatives.

correlations support a white-collar/blue-collar clustering hypothesis. The intercorrelations within the white-collar "group" are moderately large and positive. The intercorrelations within the blue-collar group are less strong but generally positive, while the intercorrelations between the two groups are moderately large and negative. The pattern of correlations for the clerical group is interesting in that this group seems to be disassociated from the white-collar group somewhat, yet not associated with the blue-collar grouping, except service workers. This latter effect may in part be due to a high predominance of non-whites in these two categories. The generally small magnitude of the correlations reflects the more uniform distribution of the clerical group by sectors and rings, as well, as evidenced by the graph of mean percentages.

A convenient method of summarizing the correlation matrix for the sub-areas is provided by the principal components loadings, generated as a by-product of the computations for the canonical correlations. The percent of variance explained by the first component ranged from 43-60% for the sub-areas examined (Rings 0-1, 2, 4, 6 and Sectors 2, 5). Only the loadings for the first three components will be reported in Table III. It can be seen that the pattern observed in the intercorrelation matrix for the region as a whole is, in general, repeated in the first component for each of the sub-areas; there appear negative loadings for blue-collar workers, less strong but negative loadings for clerical workers, and

TABLE III

Occupational Residential Factors
Principal Components Loadings
Boston Study Region and Sub-areas:1960
Census Tract as Unit

Occupation/Area	Region	Ring 0-1	Ring 2	Ring 4	Ring 6	Sector 2	Sector 5
First Component							
Professional	.86	.84	.78	.93	.89	.91	.91
Managerial	.90	.78	.85	.95	.93	.92	.91
Sales	.83	.63	.78	.82	.78	.80	.86
Clerical	-.26	.00	-.34	-.31	-.20	-.37	-.46
Skilled	-.34	-.67	-.68	-.65	.26	.14	-.76
Semi-Skilled	-.82	-.85	-.81	-.92	-.89	-.77	-.77
Service	-.48	.16	-.55	-.66	-.64	-.70	-.59
Labor	-.69	-.73	-.82	-.77	-.48	-.62	-.71
% of Trace	48%	43%	52%	60%	47%	49%	58%
Second Component							
Professional	-.16	.11	-.47	-.01	-.22	-.27	-.16
Managerial	.12	.38	.23	.15	-.09	-.16	.19
Sales	.05	.39	.43	-.14	.16	.04	-.01
Clerical	-.43	.49	-.29	-.78	.82	.42	-.73
Skilled	.70	.49	.49	-.60	.29	.90	.42
Semi-Skilled	.32	.25	.41	.11	.16	.09	.52
Service	-.73	-.85	-.64	.41	-.11	-.38	-.68
Labor	-.04	-.07	-.06	.37	-.64	-.31	.06
% of Trace	17%	20%	17%	17%	16%	17%	19%
Third Component							
Professional	-.06	-.12	-.05	-.13	.18	.04	.08
Managerial	-.09	-.17	-.02	.10	.06	-.01	.15
Sales	.14	-.16	.09	.39	.28	.28	.03
Clerical	.80	.85	.86	.49	.26	.76	-.30
Skilled	.44	-.03	.30	-.31	-.85	-.14	-.06
Semi-Skilled	-.09	-.11	-.20	-.10	-.06	-.47	-.20
Service	-.06	.29	-.18	.33	.45	.28	.22
Labor	-.40	-.23	-.27	.28	-.14	.18	.64
% of Trace	13%	12%	12%	9%	14%	13%	8%

SOURCE: Census Tract Tape

NOTE: "Percent of trace" refers to amount of variance in correlation matrix accounted for by given component. Loadings are interpretable as correlation coéfficents for variable with factor. Variables defined as in Table II.

positive loadings for other white-collar groups. This pattern weakens at the center of the region (Rings 0-1) where service workers have a slight positive loading, and in Ring 6 and Sector 2 where foremen load positively.

It would seem from the results of the percentages as well as the components, that there is a good bit of variance in the distribution of occupations within the region which might be attributed to "class" lines, roughly drawn, and that deviations from this pattern occur for economically relevant reasons.

Where transportation constraints are lifted, for example, we would expect a greater polarization of loadings in that households would be more free to cluster on a class basis than where they are hemmed in by heavier transport costs or rent. If the total outlay for these two budget items is fixed, higher rent will mean shorter distance to work or lower rent will mean longer commuting. Kain's model states that, since the rent structure declines outward from the center, those who live farther from work will be forced toward the outside of the region to balance rent and transport costs. However, if the structure of transport costs is such that annular movement is relatively facilitated, while radial movement is not, there will be a checking of outward movement and an intensification of annular movement. But, if the rent gradient is fixed within a ring, the savings on marginal transport costs may be used in the acquisition of a more suitable neighborhood. If this means similar neighbors, increased differentiation would be the expected result. In this regard, it is of some interest that Ring 4, which straddles Route 128, has the greatest polarization of loadings. If class lines were significant contributors to the variance of residential location, in addition to income and work locations, then we would expect this effect to appear most strongly where transportation constraints were lifted. This appears to be the case.

Household locations at the center of the region may be characterized as constrained largely by site-rent costs. Location in the outer-most rings, which are substantially non-urbanized, may be characterized as constrained by transport costs -- while little in rent savings may be gained from a location farther outward from the center, the opportunities to obtain housing are more scattered. A somewhat longer journey to work was found by Kain[1] for the outside ring of Detroit than for the next inner ring. While his theory would predict that a household would have little to gain from a longer journey to work, due to the flattened price curve in this area, his data suggest that adequate housing is scattered. The longer journey will cut back on rent-paying ability and constrict the household's options. For these reasons, the apparent gradient pattern of the variance explained by the first component for the sample of rings is significant. It is a pattern marked by a low amount of variance explained at the center of the region, a rise to Ring 4, and a decline to

1 Kain, op. cit., p. 145.

Ring 6. Of course, this involves interpolation, but the apparent gradient would give tentative support to our hypothesis.

The components of the residuals left from the first components showed some patterns. They are not as strong as the first, although they do indicate possible secondary clustering of upper blue-collar groups, a clustering of the lowest white-collar group and the highest blue-collar group, and a tertiary clustering of clerks, sales and service categories. These results are also in Table III. Factors two and three accounted for about 30% of the variance. One pattern in factor two was positive loadings on skilled and semi-skilled categories, seemingly associated with negative loadings on the clerical and service categories. Where this pattern was not present, factor two contained positive loadings on clerical and skilled categories, with negative loadings on laborers and, to a lesser extent, service categories. Where the foreman-operative pattern was present in factor two, the foreman-clerk pattern was shifted to factor three. Where this pattern was not present in factor three, high loadings on clerical, sales, and service were present.

Patterns such as these -- clustering of upper blue-collar, clustering of lower white-collar and upper blue-collar, and a clustering of (perhaps the lower-skilled, lower-paid) clerks, sales, and service categories -- as lower order effects contained in the observed patterns of correlations after a fairly consistent white-collar/blue-collar pairing effect has been removed are not unreasonable. However, one qualification is that we have selected patterns of loadings common to several sub-areas; the complete set of loadings for any one area reveals other occupational categories loading highly on a component in any one sub-area. While the common threads are of some interest, their reliability is a moot point.

Canonical Correlations:

In order to simultaneously determine the importance of income and education for occupational clustering, canonical correlations were computed.[1] This technique may be viewed as a generalized weighted regression, as the form is similar to a regression equation with several dependent variables. As was noted above, however, neither set of variables can be labeled "dependent." In fact, both sets consist of independent variables. The sets of coefficients obtained are used in predicting a "canonical variate" for each; the coefficients are determined in such a way as to maximize the product moment correlation between the canonical variates. This correlation is the canonical correlation, and the coefficients are interpreted as beta coefficients. The raw score coefficients, obtainable by the division of each beta by the standard deviation of its corresponding variable only (the standard deviation of the canonical variates being unity),

---

[1] Brief, non-technical descriptions may be found in

W.W. Cooley & P.R. Lohnes, <u>Multivariate Procedures for the Behavioral Sciences</u>, Wiley (New York:1962)

M.G. Kendall, <u>A Course in Multivariate Analysis</u>, Hafner (New York:1957) are not presented, as our interest is only in the relative importance of the variables. More than one set of coefficients may be obtained by repeating the computations on the residual variance after each set has been extracted. Since the smaller variable set had only two variables, no more than two functions could be computed. In some cases, this second function was not statistically significant. What we should expect from these computations is a determination of the relative importance of education and income in "predicting," in a sense, those occupational groupings which have the large coefficients.

For the region as a whole, as can be seen from Table IV, the coefficients of income and education are of equal magnitude and are of the same sign for both. If the regional pattern is conceived as an interaction between an annular rent-transportation effect and a sectorial class bias, this result is as one would expect. It is interesting that only three occupational categories, managers, skilled and semi-skilled, have even moderate-sized coefficients.

Since it was not practical to repeat the computations for all sub-regions, a sample consisting of Rings 0-1, 2, 4, and 6 and Sectors 2 and 5 was chosen. This, it was felt, covered the region fairly well, as there were rings from the center to the outside of the region represented, as well as the sector with the most differentiation of occupational distribution and one which was "representative," in a sense, of the northern and southern sectors. When this was done, education and income become more distinct in their importance, with education generally receiving the higher weighting on the first function.

The weightings for these two variables on the second function were switched, while the signs of income tended to reverse. This seems to be an artifact of the root extraction process, since there was only one "degree of freedom" in the income-education correlation matrix. Other occupational categories also took on importance.

For the first function within each sub-area, education had relatively more importance in Rings 0-1, and Ring 4, as well as for both sectors, while income was of overwhelming importance in Ring 2, and the two were of equal magnitude again in Ring 6.[1] Of course, while education has a generally higher weights, the coefficients for income are also moderate. The equal loadings for education and income in Ring 6 are similar to those for the region as a whole. In view of the relative lack of organization hypothesized for this ring, this is a comforting result. That the highest positive coefficient for education appears in Ring 4, for the rings, and this tends to support our earlier observations concerning the role of transportation in causing greater differentiation within this ring. The negative education coefficient for Ring 0-1, with its associated negative loadings on white-collar occupations, and the high income coefficient for Ring 2

[1] This illustrates one advantage of canonical correlation, for, from an inspection of the basic correlation matrix, it would appear that Ring 4 would come out similar to the region as a whole, and that Ring 6 would be similar to Sector 2, due to the general similarity of the two sets of correlation coefficients.

TABLE IV

Occupational Residential Factors
Canonical Correlations and Coefficients
Boston Study Region and Sub-areas:1960
Census Tract as Unit

Areas	Region	Ring 0-1	Ring 2	Ring 4	Ring 6	Sector 2	Sector 5
First Fucntion							
Canon-R	.87	.92	.72	.95	.94	.94	.92
R-square	.75	.84	.52	.90	.89	.88	.85
"Left-hand" Betas							
Income (Median)	.72	.38	.99	.47	.74	.36	.32
Education (Median)	.69	-.92	.13	.88	.67	.93	.95
"Rignt-hand" Betas							
Professional	.21	-.61	.32	.09	.59	.63	.48
Managerial	.70	-.03	.64	.84	.51	.28	.50
Sales	.16	.08	.17	.05	.21	.00	-.18
Clerical	.23	-.24	.18	.08	.19	.30	.09
Skilled	.47	.28	.57	.39	.52	.33	.35
Semi-skilled	-.34	.66	.22	-.25	-.09	-.51	-.59
Service	-.21	.13	.14	-.25	-.01	-.19	-.06
Labor	-.04	.19	.16	-.02	-.16	-.16	-.03

SOURCE: Census Tract Tape

NOTE: Variables defined as in Table II.

would seem to be a part of the general gradient pattern.

Turning to the occupational groupings, we see that the professional, manager, skilled and semi-skilled categories have moderate weightings for all or most of the areas. Professionals in Ring 0-1 have the same negative sign as does education, while the skilled and semi-skilled categories have reversed sign. In Ring 2, professionals, managers, and skilled workers have high importance in a function which also gives high importance to income. The skilled and semi-skilled categories have heavy weightings, with the same sign as income, suggesting that rent factors are important, as well. For Ring 4, education has a haevy weighting as do the manager and skilled categories, while the semi-skilled and service workers have negative weightings. Class lines, with skilled workers crossing them, are most apparent from the weightings in this ring. The absence of professionals from this function is apparently not due to the effective upper bound on education disturbing the fit, where the proportion of professionals rises steadily but median education stabalizes after a certain point. The regression of professionals on income and education for this ring yielded a much larger beta for education than income. Thus, particularly in the case of high opportunity for differentiation within a ring, other factors must be examined to successfully predict the distribution of professionals.

These results are far from satisfying, but they are suggestive. It may be that canonical correlation with only two variables in one set tends to place too great a restriction on the results.

The addition of several other variables to the income-education set was also made. In general, these variables were chosen to represent external aspects of a neighborhood -- aspects which would serve as the setting for a location decision. These variables were median rooms per dwelling, and percentages of foreign stock, structure age 20+ years, single family dwelling units, rental status, and length of residence 20+ years. Computations were carried out only for the region as a whole (see Table V).

Again, the median income and median years of education variables had large coefficients on the first function. Single family dwellings was also moderate and foreign stock was of more importance than the rest. Again, the managerial, skilled and semi-skilled categories have heavier weights than the others, although the professional category also appeared.

The importance of the education variable in the first two functions indicates the importance of life style variables in residential clustering, even with the effects of others considered. Income is important in the first and third functions and does not reappear until the last function (not showen). On the other hand, length of residence 20+ years and structure age 20+ years are relatively unimportant throughout, while foreign stock and median rooms are of moderate importance. Rental status and single family dwelling units are the only two additions of fairly consistent importance.

For the occupations, managerial, skilled and semi--skilled categories are the strongest overall, while the

TABLE V

Occupational Residential Factors
Canonical Correlations and Coefficients
Contextual Variables
Boston Study Region: 1960
Census Tract as Unit

Functions	I	II	III	IV	V
Canon-R	.91	.72	.56	.41	.27
R-square	.83	.51	.31	.17	.07
"Left-hand" Betas					
Income (Median)	.48	.07	.46	.01	.21
Education (Median)	.64	.70	-.15	.09	-.09
Rooms (Median)	.13	-.01	.50	-.30	-.09
Foreign Stock	.25	-.06	.20	.11	.35
Struct. Age 20+ Yrs.	.10	-.02	-.09	.07	.22
Rental Status	.01	.69	.53	-.78	.44
Residence 20+ Yrs.	-.17	-.02	.37	.02	-.30
Single Family DU	.49	-.17	-.24	-.54	.69
"Right-hand" Betas					
Professional	.21	.18	.10	.11	.02
Managerial	.77	-.60	.38	-.11	-.03
Sales	.12	.11	.37	.36	.69
Clerical	.10	.38	.50	.08	-.29
Skilled	.51	-.47	.00	.64	-.17
Semi-skilled8	-.24	-.45	.64	-.06	.45
Service	-.14	-.03	.03	.65	.44
Labor	-.04	-.16	.23	-.07	-.06

SOURCE: Census Tract Tape

NOTE: Occupation variables defined as in Table II. Foreign Stock is percentage of total population. Structure Age, Rental Status, Residence, and Single Family DU are percentages of total housing units.

clerical category emerges with some strength after the first function, and sales after the second. Those at the top, the professionals, and at the bottom, the laborers, do not seem to be well indexed at all. While this is surprising in that we would expect these two to have the most marked amount of clustering, in terms of the above discussion, the variables chosen as predictors do not reflect this.

The addition of these additional variables does not seem to materially improve the interpretation. The exception is the additional importance of rental status and single family dwelling, which seem to replace income, to a degree. Apparently, education is not a proxy for any of the other variables considered here.

Components Analysis

The "effect of aggregation" is one troublesome problem in the analysis of areal data. Two conflicting problems are apparent. One is that if large units are employed, the within-area variation is likely to rise, possibly at the expense of the between-area variation. Since the smaller units are aggregated, their effects may cancel each other out, producing a uniform shade of gray. At the least, clear-cut distinctions may blur somewhat. The town of Newton is a good example of this, with one tract of quite low "status" based on income, education, and occupation, while most other tracts are fairly high. On the other hand, it is also obvious in general that, if one small sub-area has a certain set of characteristics, it is more likely that adjoining areas will be similar.

This lack of independence is troublesome, and is similar to the problems of serial and autocorrelation in economic time series. In fact, Geary has proposed a "contiguity ratio"[1] based on von Neumann's ratio of mean square successive differences to the variance.[2] Since an area may have a large number of other areas adjacent to it the ratio proposed by von Neumann is not appropriate. Instead, Geary proposes a generalization of the von Neumann ratio to higher dimensions. Geary supplies standard errors for testing this under the assumptions of a) complete sampling of the universe, and b) a random sample from a normal universe, for testing the significance of the ratio.

---

[1] R.C. Geary, "The Contiguity Ratio and Statistical Mapping," The Incorporated Statistician: 5 (1954), 115-145.

[2] G. Tintner, Econometrics, Wiley (New York:1965), 252-255.

The purpose of this discussion is not as a prelude to the use of the technique, but only to indicate that the problems of dealing with economic time series and areal data are similar, and to indicate that too great a concern with small, homogeneous areal units may be unfounded. Aggregation may even provide a solution to some of these problems, as well as reducing the amount of computation. The Duncans, for example, found that their index of dissimilarity computed for the 1,178 census tracts of the Chicago area correlated .98 with that same index computed for 104 aggregate sector-zone segments.[1]

In order to see what effect aggregation might have on factor analysis, a components analysis was performed on a set of variables from census tract data which was comparable to a components analysis of data at the town level.[2] Several problems of comparability, aside from simple aggregation, are discussed below.

Comparability of Samples:

Thirty-five variables were included in the analysis by town and consisted of the percentage of households in the town which were included in various categories of dwelling unit type, persons per dwelling unit, length of tenure, household income, and rental status, as well as information on the percentage of persons falling into categories of age, occupation,

---

1 Duncan and Duncan, op. cit.

2 Clarke, op. cit.

and industry type. The data were obtained from the results of a transportation study of the Boston region, resulting in a file of about 40,000 household interviews and about 120,000 individual interviews. This represented a sample of from 3-7% of the households, depending upon the section of the region. A lower sampling percentage held in the center of the region, while a higher percentage held outside Route 128 (Rings 5-7). This seems to have been done in order to assure that the sample size was adequate in the less dense areas, but the sampling percentage also varied within these broad areas as well, so that the weighting of the households was necessary in order to obtain estimates of the proportions. In addition, the data on persons was weighted by the dwelling unit factor, since information was gathered on all persons over 5 years of age in the household -- this next stage of the sampling can thus essentially be ignored for the purposes of estimates. Any necessary weighting and adjusting of the tract data was done by the Census -- of course, the Census figures are the criterion against which all other surveys of this type must be measured.

Aside from sampling variability, the actual times of the two samples are different, which will add an additional amount of variability. The time of the Census enumeration was 1959, while the traffic survey was conducted in late 1963 and early 1964. Overlooking the variability inherent within the extended time-span of the traffic survey itself, almost 5 years separates the two surveys -- certainly enough time for a considerable amount of readjustment to take place.

Finally, there are differences in a few of the variables themselves. The percentage of households with two or more persons employed was used in the analysis by towns, while the total employed over the total population was used for the tracts. Due to excessive missing observations in the 5+ dwelling-unit structure-type information for the tracts, the percent of households in 3+ dwelling-unit structures was used. There was no breakdown by number of persons per dwelling for the tracts (four categories for the towns), so a crude average of persons over households was substituted. The number of persons ages 20+ was missing from the tract tape. Finally, there were differences in industry classifications, but these were not as pronounced as some of the other discrepancies. These defects do not make this a good test of only the "effects of aggregation," but the results are interesting, nonetheless. A better comparison would be to carry out parallel component analysis on the tract data itself and on the same data aggregated to the town level, or higher, but, as the first half of the comparison was already in existence, it seemed more reasonable to capitalize on it rather than to undertake another lengthy series of computations.

Principal components analysis was used on the town data, so it was used on the tract data as well. Thus, all of the variance -- that attributable to these variables in general, to these variables in this region, and to error of several sorts -- is included in the resulting factors. This is not a wholly desirable situation; the alternatives are such that its use is not indefensible, either.

For example, if only the given variables and given observations are of interest, communality estimates could be subsituted for the 1's in the main diagonal of the correlation matrix. These might be the squared multiple correlation of all other variables against each variable in turn. If this is done, the off-diagonal elements of the r-matrix should be adjusted as well, to further remove the variance unique to individual variables. The reasonableness of this procedure breaks down if the variables and observations do not define the domain of interest, however. In addition, a sampling theory for the loadings has not, to our knowledge, been formulated, making the choice of the "significance" of a loading arbitrary, and comparisons impressionistic.

Other approaches are possible -- the maximum likelihood method of Lawley, for example[1], or a procedure which seeks to maximize the canonical correlation between the "raw" variables and that part of the same score with the uniqueness of each variable removed. This obviously heavily weights variables with high communality, so as to make the correlation between the two sets of scores a maximum. These two procedures involve a good deal more computational time, with the existing programs, than do the previous two, and were not considered practical.

---

[1] D.N. Lawley and A.E. Maxwell, Factor Analysis as a Statistical Method, Butterworths (London:1963), Chapter 2.

Results:

The comparison of the components analysis of the town data (Clarke) with the tract components (TableVI) shows many similarities. The sample of tracts tended to accentuate the pattern of loadings. Those with high loadings, particularly on the first component, tended to be strengthened, while those which were low tended to be weakened. Reversals of sign occurred for some variables which had low loadings on a component; this was most common for the weaker components, and for loadings which were below .3, giving some indication of what the standard error of the loadings may be for a sample of 150. In the town sample, the first component consisted of negative loadings on variables representing "familism" -- single family housing units, ages 5-15, 30-50, school students -- and $7000+ income. Positive loadings occurred on complementary variables such as smaller household size, multi-family dwellings, age 16-29, rental status, and low income. This was repeated at the tract level, with the addition of several industry, occupational, and income categories.

The most important differences were the addition of moderate loadings on occupation professional and skilled, and percent resident 0-5 years. The changes in the occupational categories served to bring out a white-collar/blue-collar split, which had not been apparent in the analysis by town. Professionals changed from a low loading on the "suburban" side (positive loadings) to a moderate loading on the "urbanism" side as did percent resident 0-5 years, while skilled workers has a moderate loading on the "suburbanism" side. This

TABLE VI

Principal Components by Tract

PRINCIPAL COMPONENTS FACTOR LOADINGS

VARIABLE DESCRIPTION		VAR NO.	1	2	3	4	COMMUNALITY
Single Family DU	1	VAR 1	-.705	-.534	-.109	-.038	.796
Multi-Family DU	2	VAR 2	.743	.504	-.131	-.076	.830
Pers Per Hshld	3	VAR 3	-.205	.013	-.042	.119	.058
Age 5-9	4	VAR 4	-.863	-.025	-.309	-.194	.878
Age 10-14	5	VAR 5	-.873	.006	-.206	-.111	.817
Age 15-19	6	VAR 6	.185	-.122	-.326	.698	.642
School students	7	VAR 7	-.875	-.058	-.202	-.142	.829
College students	8	VAR 8	.515	-.326	-.230	.609	.796
Employed % of Pot	9	VAR 9	.701	-.077	.491	.015	.738
Occ Professional	10	VAR 10	.357	-.822	-.009	.110	.816
Occ Sales + Clerk	11	VAR 11	.400	-.456	.648	-.013	.788
Occ Skilled	12	VAR 12	-.530	.267	.532	.001	.635
Occ Semi-skilled	13	VAR 13	-.100	.870	.139	.175	.817
Occ Unskilled	14	VAR 14	.045	.652	-.035	-.089	.436
Occ Service	15	VAR 15	.641	.226	-.041	-.325	.569
Indy Retail	16	VAR 16	.355	.012	.571	-.377	.594
Indy Construction	17	VAR 17	-.470	.112	.279	-.140	.331
Indy Manufacturing	18	VAR 18	-.318	.590	.264	.244	.579
Indy Wholesale	19	VAR 19	.245	-.428	.267	-.283	.395
Indy Professional	20	VAR 20	.624	-.604	-.061	.216	.804
Indy Pers. Service	21	VAR 21	.572	-.142	-.115	-.461	.573
Indy Government	22	VAR 22	.136	-.121	.510	.003	.293
Rented DU	23	VAR 23	.760	.538	-.019	.037	.868
Here 0-5 Yrs	24	VAR 24	.400	-.025	-.349	-.244	.342
Here 20+ Yrs	25	VAR 25	-.196	.157	.541	.383	.502
Income $0-3999	26	VAR 26	.418	.717	-.414	-.097	.870
Income $4-6999	27	VAR 27	-.193	.689	.362	.188	.678
Income $7-9999	28	VAR 28	-.485	-.458	.466	.058	.666
Income $10,000 +	29	VAR 29	-.019	-.883	-.087	-.059	.790
		LATENT ROOTS	7.613	6.095	3.098	1.926	18.731

would indicate somewhat more strongly that this component is a housing type factor.

The loadings on component two tend in the direction of a white-collar/blue-collar dichotomy for the towns -- strong negative loadings on professional and income $10,000+ and positive loadings on skilled and semi-skilled occupations and income $4-7000. At the tract level, this is strengthened by the addition of higher negative loadings on single-family dwellings, occupation clerk-sales, and income $7-10,000, as well as stronger positive loadings on multi-family units, rental status and low income. These additions were in the nature of raising a loading from below .3 to above .5, a point where it is no doubt non-zero. In this case, these additions served to clarify a social rank relationship which could only be guessed at from the results by town.

The third component had two significant additions, at the tract level. Initially consisting of a high proportion of families residing 20+ years and salesmen and a negative relation to residence 0-5 years, the tract analysis made the significant (in terms of the previous discussion) addition of foremen and middle-income ranges, and thus better represented a secondary clustering of lower white-collar and upper blue-collar workers -- a pattern which was not apparent at the town level. The income relation mirrors the occupational categories. The fourth component, with slight positive loadings on residence 0-5 years and salesmen (for towns) had the sign of the residence 0-5 years loading changed, for tracts. The extraction of roots was stopped after four.

Another run was made using dummy variables for sectors and rings (Table VII) since the factor scores had been mapped for towns, and the use of dummies would allow a partial comparison to be made without the necessity for mapping all of the tracts' scores.[1] The use of dummy variables allows the introduction of information which cannot be scaled or is inappropriate for certain other reasons, such as curvilinearity.[2] In regression analysis, where this was first developed, the effect of region of the country, of sex, or of ethnicity might be important to the analysis and the use of dummy variables would be one way of introducing it. A discriminant analysis may be viewed as a regression with a dummy dependent variable, as well. The dummy variable itself has the value one if the quality is present and zero if it is absent. For example, the variable "Sector 5" was scored one if the tract was located in Sector 5; it was scored zero otherwise. As the rings are ordered by distance from the center of the region, one variable, instead of six, could have been used for rings, but it would not have been meaningful to do this for the sectors. The addition of the dummy variables did not materially alter the loadings for the other variables on the first three components. A fifth component was added, due to the use of a different stopping criterion, but it was made

---

[1] See Clarke's "Static Factor" maps, Clarke, op. cit.

[2] See D.B. Suits, "Use of Dummy Variables in Regression Equations," J. Am. Stat. Assoc.: 52 (1957), pp. 548-551. J.N. Morgan, et al., Income and Welfare in the United States, McGraw-Hill (New York:1962), make extensive use of this technique in their analysis.

TABLE XII

Principal Components by Tract With Regional Dummies

PRINCIPAL COMPONENTS FACTOR LOADINGS

VARIABLE DESCRIPTION		VAR	NO.	1	2	3	4	5	COMMUNALITY
SINGLE FAMILY	1	VAR	1	-.753	-.468	-.173	-.070	-.233	.876
MULTI-FAMILY	2	VAR	2	.788	.449	-.086	.143	.114	.863
Pers Per Hshold	3	VAR	3	-.196	.026	-.050	-.161	-.081	.074
Age 5-9	4	VAR	4	-.844	.049	-.318	.235	.042	.874
Age 10-14	5	VAR	5	-.853	.083	-.198	.191	.180	.844
Age 15-19	6	VAR	6	.182	-.125	-.275	-.341	.579	.576
School Students	7	VAR	7	-.858	.018	-.199	.234	.168	.859
College Students	8	VAR	8	.484	-.356	-.215	-.436	.305	.690
Employed % of Tot	9	VAR	9	.674	-.147	.451	-.226	-.344	.849
Occ Professional	10	VAR	10	.284	-.836	-.059	-.247	-.126	.860
Occ Sales + Clerk	11	VAR	11	.362	-.508	.657	.127	.141	.857
Occ skilled	12	VAR	12	-.515	.285	.519	.006	-.178	.648
Occ Semi-skilled	13	VAR	13	-.046	.862	.161	-.269	-.086	.851
Occ Unskilled	14	VAR	14	.104	.656	-.041	.107	-.022	.455
Occ Service	15	VAR	15	.658	.181	-.074	.213	-.327	.624
Indy Retail	16	VAR	16	.337	-.040	.559	.306	-.301	.612
Indy Construction	17	VAR	17	-.457	.129	.252	.152	-.184	.346
Indy Manufacturing	18	VAR	18	-.299	.592	.262	-.476	-.255	.800
Indy Wholesale	19	VAR	19	.233	-.457	.271	.355	.119	.476
Indy Professional	20	VAR	20	.573	-.640	-.096	-.267	-.020	.819
Indy Pers Service	21	VAR	21	.558	-.179	-.156	.234	-.462	.636
Indy Government	22	VAR	22	.139	-.132	.516	.226	.363	.486
Rented DU	23	VAR	23	.805	.471	.039	.014	.171	.901
Here 0-5 Yrs	24	VAR	24	.397	-.057	-.339	.186	-.150	.333
Here 20+ Yrs	25	VAR	25	-.183	.156	.554	-.233	.325	.524
Income $0-3999	26	VAR	26	.475	.699	-.367	.084	.027	.856
Income $4-6999	27	VAR	27	-.140	.686	.384	-.063	.072	.646
Income $7-9999	28	VAR	28	-.529	-.432	.419	-.053	-.097	.654
Income $10,000 +	29	VAR	29	-.083	-.877	-.121	-.010	-.021	.791
RING 0-1		VAR	30	.482	.378	-.260	.154	.061	.471
RING 2		VAR	31	.362	-.066	.294	.190	.304	.350
RING 3		VAR	32	-.112	-.277	.282	.008	.082	.175
RING 4		VAR	33	-.218	-.227	-.145	.071	-.058	.128
RING 5		VAR	34	-.268	-.110	-.152	.189	-.206	.185
RING 6		VAR	35	-.232	.199	-.049	-.450	-.225	.349
SECTOR 1		VAR	36	-.100	.150	.087	-.065	-.085	.052
SECTOR 2		VAR	37	-.163	.144	.074	-.268	-.010	.125
SECTOR 4		VAR	38	.130	-.134	-.127	-.360	.130	.198
SECTOR 5		VAR	39	.181	-.407	-.039	-.001	-.213	.245
SECTOR 6		VAR	40	.168	.135	-.230	.307	-.039	.195
SECTOR 7-8		VAR	41	-.123	-.009	.137	.428	.290	.301
		LATENT ROOTS		8.223	6.614	3.378	2.297	1.941	22.453

up of miscellaneous loadings. Neither component four nor five will be discussed, due to their lack of comparability with the analysis by towns.

On factor one, the urbanism factor, Rings 0-1 and 2 load positively, to a degree, while the Rings 4-6 load negatively, The sectoral variation is generally low. This is also the pattern as mapped by Clarke; a concentration of "low extreme" values in Rings 4-6 and "high extreme" values in Rings 0-2. The white-collar/blue-collar factor 2 follows the map fairly well. A positive loading is given to Ring 0-1, while negative loadings appear in Rings 3 and 4, both of which contain a fair number of "low extreme" factor scores in Clarke's maps. The loading becomes positive again in Ring 6, where there are a moderate number of "high extremes" in the maps. The sectoral pattern is strong for this factor with many of the low extreme scores falling in Sector 5, and many others in Sector 4, along with some high extremes. The signs of the loadings on the dummies are negative, and the magnitudes are ordered correctly, in view of the mixing of high and low extremes in Sector 4. For the clerical-foreman factor 3, a negative loading is found for Ring 0-1, and positive loadings for Rings 2 and 3; this corresponds with indications from the map. With the exception of Sector 6, the other sub-areas have generally low loadings. That Sector 6's negative loading does not correspond to the mixed indications from the map is no doubt due in part to the mapping of only extremesvalues.

This brief comparison of the maps and the dummy variate loadings indicates that such an approach may be of some use in lieu of actual mapping. To cover the region well would require many more dummy variables; in the extreme case, one would be required for each sector-ring intersection, but this would be totally unreasonable. To get such a complete picture would require actual mapping, but the dummy variable technique can be of some preliminary value.

In general, the comparison of the factor analyses has shown that the town may be too large a unit of observation to use in statistical work of this sort, although such a conclusion is to be qualified in several ways. First, there is the lack of strict comparability between data sources and variable definitions. Also, there is no good criterion with which to measure the success or failure of a factor analysis. In comparing the results of identical analyses on independent populations, Lawley and Maxwell's[1] technique may be applied. The advantage of this technique is that tests of significance for similarity are provided. We were limited to simply evaluating the meaningfulness of the results obtained; while this is open to questions of validity, the analysis by tract does seem to have yielded loadings which are as meaningful or more meaningful than the analysis by town.

In the first three components, the basic interpretation remains unchanged. The first was essentially an urbanism-

---

1. Lawley and Maxwell, op. cit., Chapter 8.

suburbanism factor, the second a social rank factor, and the third apparently a clustering of upper blue-collar and lower white-collar workers. By this test, it would appear that the town level can give rough characterizations, but that these are sharpened at the tract level.

Chapter IV

SUMMARY AND CONCLUSIONS

We have examined both literature and data, principally concerning household location, and have noted that economic factors would seem to give a region its overall structure, while social factors operate within these constraints. In addition, several other points stood out.

1. The high-status wedge in the western part of the region, as found by Clarke, was quite apparent here, as well.
2. We found that class lines, most strongly represented by a white-collar/blue-collar split, emerged in all of the results, as one strong factor in differentiation.
3. The amount of this differentiation seemed to form a gradient pattern, with its peak at a ring served by a high-speed, circumfrential highway, indicating one role of transportation systems is that of aiding in the differentiation of the population.
4. Median education, our proxy variable for life style, performed, on the whole, better than the median income variable in indexing occupational clustering.
5. Our components factor analysis of tract data yielded results which were somewhat better defined than an analysis of data at the town level. There were more similarities than differences between the two, indicating that the town level is not too gross a level for statistical analysis of this type, but other differences indicate that conceptually clearer results are obtained at the tract level.

We asked at the outset what value the study of household and migration patterns could have for planning. The imperfections of the market and the penalties paid by a region's inhabitants are brought out clearly. One penalty, for example, which is incurred by non-whites in a ghetto is clear enough; for non-centrally employed non-whites the separation from workplace may be much greater than if they were free to settle at will, as has been pointed out many times. What is often not mentioned is the costs to lower-income, centrally-employed whites, since they, too, may be farther removed from their workplace than they ordinarily would be if central city land was not pre-empted.

When renewal work and building activity in the central city is examined in the context of population movements in and into the region, an interesting pattern imerges. Several aspects are well known. Radial transport routes are, of course, being strengthened, providing good access to large amounts of undeveloped land in the outer rings. Per capita income has been on the rise in the region, in part due to an efflux of lower-paid workers and the influx of higher-paid workers and better salaries for residents as a part of the general upgrading of the region's economic base. The effects of these factors may be characterized, in terms of the location theory treated, as leading to increased expansion of residences within the region; the flattening of the curves representing the price structure and bid-rent responces of households will accomplish the spread.

One aspect of the population shifts is quite relevant here; this is the emigration from the region of individuals 25-35 years of age. Its effect in softening the suburban home market is obvious, as a large proportion of home buyers come from this age group. At the next remove, though, the effect upon the multi-family dwelling market should also be considered. In particular, will there be a softening of this market in the next decade. The answer would seem to be in the affirmative since present construction is encouraged by a presently favorable age distribution. As the cohort of older apartment dwellers is thinned out, the next cohort may not be able to take up the slack, due to an inability to sell their present home. With an unfavorable balance of younger persons available for multi-family units, as well, the prospects would not seem encouraging.

A substantial amount of redevelopment work, including multi-family units, is taking place in the center of Boston; according to current reports, this work will begin to show profit only in the next decade. A further decline of land values in the core would endanger the current work, as well as hopes for the future. One remedy for such a situation might be maximum-lot zoning, in an effort to shift the desirability of centrally-located land back towards the center. Such a plan would not be suitable to outlying towns, however, and would have to be imposed from above. In the least, though, it would seem desirable to examine this in detail as suggestive of the direction Boston might go in safeguarding its investment.

Appendix

FURTHER NOTES ON DATA AND METHODS

Data and Variable Definition

The data for Boston-region Census tracts was contained on a magnetic tape obtained from the Joint Center for Urban Studies. As this tape was in Fortran-IV binary format, while the programs which would use it were to be run under the Fortran-II Moniter System, a new tape had to be prepared which was compatible with the input routines of this latter system. Two approaches were possible. On the one hand, it would have been possible to write a Fortran-IV program to punch the data onto cards, or write a tape in BCD, card-image format, and to subsequently feed this into the statistical programs. The other approach was to rewrite the binary tape directly into "FMS-binary" form. Due to the relative slowness of BCD-format conversion, a wise procedure would have been to rewrite the card-image data into a binary form again in the first stage of the statistical processing. In view of this, it seemed most reasonable to make the Fortran-IV - to - Fortran-II conversion of the binary tape.

The existence of a variable-buffer-length binary read routine as a part of the Planning Department's SYSTEM-2 tabulation program greatly facilitated the rewriting. The rewriting was accomplished by simply reading the Fortran-IV tape with this routine, then writing the Fortran-II version with calls to the FMS output routines. At the same time, variable labels and control words were added to allow direct reading by the Data-Text System[1], which was utilized in parts of the analysis. In addition, the sector-ring

---

[1] See The Data-Text System, Preliminary Manual (Cambridge: 1967), pp. 169-171 for a discussion of these requirements.

codes and the "Foreign Stock" item of the census were added to the data for each tract.

As the data were in the form of counts, various runs were done to convert it to percentage form, clean and check it. A new tape of percentages was produced in the Data-Text binary format by Data-Text itself. It was this tape which was used in subsequent analysis. The principal components used in the comparison with the results by town, done by Clarke, were computed using Data-Text. For the canonical correlations for the region and sub-areas, correlation matrices were computed for each area and punched using Data-Text. These correlations were then fed into the Multivariate Statistical Analyser[1], which computed the canonical correlations themselves and the components given in Table III.

While the mechanics of running these programs at MIT are not pertinent to this appendix, it should perhaps be noted that, while both were written under the Harvard FMS, both run at MIT as "non-standard FMS" jobs and both require small programs, obtainable from the authors, which perform the searches of the tapes on which the computational programs are stored. In addition, MSA requires a special FMS tape (MIT'S S7A system) for proper execution.

The actual definition of the variables has been treated in the text and tables, but can be summarized. The following tables give this summary, listing whether the variable was

---

[1] See K.J. Jones, The Multivariate Statistical Analyser, (Cambridge:1964), pp. 86-100.

a median or, if percentage, the denominator employed. One table gives the definitions for the canonical correlations, the other for the components.

TABLE A-1
Variable Definition for Canonical Correlations

VARIABLE	MEDIAN	TRACT POPULATION AS DENOMINATOR Male Labor Force	Total Pop.	Total Hsng. Units
All Occupations		X		
Income (Families and Unrelated)	X			
Education (School yrs., pers 25+)	X			
Rooms in Structure	X			
Foreign Stock			X	
Structure Age 20+				X
Rental Status				X
Residence 20+ Yrs.				X
Single Family DU				X

TABLE A-2
Variable Definition for Principal Components

VARIABLE	MEDIAN	TRACT POPULATION AS DENOMINATOR Total Pop.	Total Families	Total Hsng. Units
Single Family DU				X
3+ Family DU				X
Pop. Per Household				X
Age 5-9		X		
Age 10-14		X		
School Students		X		
College Students		X		
Employed Labor Force (Male & Female)		X		
Occ. Clerk-Sales		X		
Occ. Skilled		X		
Occ. Semi-Skilled		X		
Occ. Unskilled		X		
Occ. Service		X		
Ind. Retail		X		
Ind. Construction		X		
Ind. Manufacturing		X		
Ind. Wholesale		X		
Ind. Professional		X		
Ind. Personal Service		X		
Ind. Government		X		
Rental Status				X
Residence 0-5 Years				X

TABLE A-2 (Continued)

VARIABLE	MEDIAN	TRACT POPULATION AS DENOMINATOR Total Pop.	Total Families	Total Hsng. Units
Residence 20+ Years				X
Income $0-$3999			X	
Income $4000-$6999			X	
Income $7000-$9999			X	
Income $10,000+			X	

Statistical Techniques

Since canonical correlation is a relatively little-used technique, we will discuss it briefly, as an amplification of the text discussion, as well as indicate its relation to principal components analysis. The discussion will of necessity be rather superficial, in the interest of brevity, but should serve to give a general idea of the usage of the technique.

Whereas the multiple correlation coefficient may be thought of as the correlation of the dependent variable with the estimate of that variable obtained by weighting the independent variables with a certain set of coefficients, the canonical correlation is the correlation of two predicted dependent variables with each other. That is, given two sets of variables which are to be related, the method of solution selects two sets of coefficients, one for each set of variables, which predict two variables which are maximally correlated. The coefficients of each set indicate the degree to which each variable contributes to this relation. As it turned out in our runs, only a few variables in the occupation set contributed consistently to the relation between the two sets of variables.

Expressed in matrix notation, what is desired is a vector of coefficients for the one set used in predicting an estimated variable in the form $\hat{\underline{x}} = \underline{a}\ \underline{X}$, where $\hat{\underline{x}}$ is the (1xN) vector of estimated values, $\underline{a}$ the (1xP) vector of coefficients, and $\underline{X}$ the (PxN) data matrix (N is the number of observations and P is the number of variables). A corresponding vector $\underline{b}$ predicts the variable from the other set $\hat{\underline{y}} = \underline{b}\ \underline{Y}$. (Note that our coefficient vectors were in standard-score form, not raw-score form as we have here.) The correlation coefficient then represents the degree of relation between the $\hat{\underline{x}}$ and $\hat{\underline{y}}$ vectors of estimated scores, although the solution method does not predict these vectors and then correlate them.

Rather, by partitioning the correlation matrix of all the variables so that $\underline{R} = \begin{bmatrix} \underline{R}_{11} & \underline{R}_{12} \\ \underline{R}_{12} & \underline{R}_{22} \end{bmatrix}$, where $\underline{R}_{11}$ and $\underline{R}_{22}$ are the intercorrelations within the "right and left-hand" variable sets, while $\underline{R}_{12} = \underline{R}'_{21}$ are the intercorrelations between variables in the two sets. The equation $(\underline{R}_{22}^{-1}\ \underline{R}_{21}\ \underline{R}_{11}^{-1}\ \underline{R}_{12} - v\underline{I})\beta = \underline{0}$ is solved after the imposition of additional constraints to insure a unique solution. $\underline{I}$ is the identity matrix, v is the characteristic root of the equation, and β is its associated vector. Given β, α follows directly as $\alpha = (\underline{R}_{11}^{-1}\ \underline{R}_{12}\ \beta)/\sqrt{v}$. The root is the square of the canonical correlation. The raw score coefficient vectors $\underline{a}$ and $\underline{b}$ also follow directly from α and β.

If more than one pair of functions is desired, the process is repeated on the residuals of the R-matrix, or that portion unexplained by the first pair of functions. This process may be repeated until there are as many pairs of functions as there are variables in the smaller set or until a test of significance indicates that little is to be gained from further functions in terms of "variance explained."

Horst[1] has generalized this procedure so that more than two sets of variables may be analysed simultaneously.

For example, we might have occupation, social rank and urbanism sets, and fit sets of coefficients to these groups of variables simultaneously. The resulting weights would be expected to show which variables in the three sets had the strongest association. As programs for the calculation of this are not completed, the technique was not used.

Various methods of "factor analysis," on the other hand, attempt to derive vectors of coefficients which maximize the amount of variance explained in a dispersion matrix, or the standardized, correlation, matrix. Some of the various types of factor analysis were mentioned in the text of the report. The method used, principal components analysis, treats all of the variance found, without attempting to make adjustments for error. That is, components analysis assumes that the observed variable x_i can be reproduced entirely if P components (P is also the number of variables) are found. In matrix notation, this would be $x_i = \sum_{r=1}^{p} l_{ir} z_r$, or $\underline{x} = \underline{L}\ \underline{z}$, where l_{ir} is the loading and z_r is the component. Factor analysis proper postulates that $x_i = \sum_{r=1}^{k} l_{ir} f_r + e_i$, or $\underline{x} = \underline{L}\ \underline{f} + \underline{e}$, where f_r is a factor, e_i is the unique error associated with variable i, and $k < p$. The solution for the components can be represented by $(\underline{R} - v\underline{I})\ \underline{l} = \underline{0}$. The formal similarity of this equation to that for canonical correlation is apparent, and has been treated by Tintner.[1] The vector $\underline{l}$, when normalized, gives the loadings

---

1 P. Horst, "Relations Among n sets of Measures," Psychometrica: 26 (1961), 129-149.

1 G. Tintner, "Some Formal Relations in Multivariate Analysis," JRoy. Stat. Soc.: B-12 (1950), 95-101.

while v gives the variance explained. This procedure is repeated on the matrix $\underline{R}' = \underline{R} - \underline{l}\underline{l}'$ (where $\underline{l}\underline{l}'$ represents theouter product of the vector). $\underline{R}'$ represents the residual variance not explained by that component (and all proceeding ones). If all components of $\underline{R}$ are extracted, then $\underline{R} = \underline{L}\ \underline{L}'$. When fewer components are extracted. than there are original variables, $\underline{R} \neq \underline{L}\ \underline{L}'$. This is the case for our runs. The right-most column, labeled "communality" in Tables VI and VII is simply the principal diagonal of $\underline{L}\ \underline{L}'$, and gives the proportion of variance of that variable explained by the loadings vectors. The latent roots may be obtained from the diagonal of $\Lambda = \underline{L}'\ \underline{L}$. These are the last rows of our Tables VI & VII labeled "latent roots" and indicate the amount of variance of the entire set of variables accounted for by each component, when taken as a percent of the sum of all roots.

Our comments in the text on the "reality" of the results obtained from these techniques need not be repeated here except to say that the meaningfulness of the results is not guaranteed by such a manipulation, and that the imputation of meaning is a subjective matter. However, the results may be suggestive of avenues for further research.

List of Towns by Geographic Area

TOWN	RING	SECTOR
Abington	5	7
Acton	6	4
Amesbury	7	2
Andover	5-6	2
Arlington	2-3	3
Ashland	6	5
Avon	5	7
Ayer	7	4
Bedford	5	3
Bellingham	7	6
Belmont	3	4
Berlin	7	4
Beverly	5	1
Billerica	5-6	3
Blackstone	7	6
Bolton	7	4
Boston		
Brighton	2	5
Charlestown	1	2
E. Boston	1	1
Jamaica Plain	2	6
Hyde Park	2	6
N. Dorchester	1-2	7
Roslindale	2-3	6
Roxbury	1	6
S. Boston	1	7
S. Dorchester	2-3	6
W. Roxbury	3	6
Boxborough	6	4
Boxford	6	2
Braintree	4	7
Bridgewater	7	7
Brockton	5	7
Brookline	2-3	5
Burlington	4	3
Cambridge	1-2	4
Canton	4	6
Carlisle	5	3
Chelmsford	6	3
Chelsea	1-2	1
Clinton	7	2
Cohasset	5	8
Concord	5	4
Danvers	5	1
Dedham	4	6
Dover	4	5
Dracut	6	3
Dunstable	7	3
Duxbury	7	8
E. Bridgewater	6	7
Easton	6	7
Essex	6	1
Everett	1-2	2
Foxborough	6	6
Framingham	5-6	5
Franklin	7	6
Georgetown	7	2
Gloucester	7	1
Groton	7	3
Groveland	7	2
Halifax	7	8
Hamilton	6	1
Hanover	5	8
Hanson	6	8
Harvard	7	4
Haverhill	7	2
Hingham	4	8
Holbrook	5	7
Holliston	6	5
Hopedale	7	5
Hopkinton	7	5
Hudson	6	4
Hull	4	8
Ipswich	6	1
Lancaster	7	4
Lawrence	6	2
Lexington	4	3
Lincoln	4	4
Littleton	6	4
Lowell	6	3
Lynn	3	1
Lynfield	4	2
Malden	2	2
Manchester	6	1
Marblehead	4	1
Marlboro	6	4
Marshfield	6	8
Maynard	6	4

TOWN	RING	SECTOR
Medford	2-3	3
Medfield	5	6
Medway	6	5
Melrose	3	2
Mendon	7	5
Merrimac	7	2
Methuen	6	2
Middleton	5	2
Milford	7	5
Millis	6	6
Milton	3	7
Millville	7	5
Nahant	3	1
Natick	5	5
Needham	3	5
Newbury	7	1
Newburyport	7	1
Newton	3	5
Norfolk	6	6
N. Andover	6	2
Northborough	7	4
Northbridge	7	5
N. Reading	5	2
Norwell	5	8
Norwood	4	6
Peabody	4	1
Pembroke	6	8
Pepperell	7	3
Plainville	7	6
Quincy	3	7
Randolph	4	7
Reading	4	2
Revere	3	1
Rockland	5	8
Rockport	7	1
Rowley	6	1
Salem	4	1
Salisbury	7	1
Saugus	3	1-2
Scituate	5	8
Sharon	5	6
Sherborn	5	5
Shirley	7	4
Somerville	1-2	3
Southborough	6	5
Stoneham	3	2
Stoughton	5	7
Stow	6	4

TOWN	RING	SECTOR
Sudbury	5	4
Swampscot	4	4
Tewksbury	5-6	3
Topsfield	6	1
Tyngsborough	7	3
Upton	7	5
Uxbridge	7	5
Wakefield	3-4	2
Walpole	5	7
Waltham	3-4	4
Watertown	2-3	4
Wayland	5	4
Wenham	5	1
Wellesley	4	5
Westborough	7	5
W. Bridgewater	6	7
Westford	6	3
Weston	4	4
W. Newbury	7	2
Westwood	4	6
Weymouth	4	8
Whitman	6	7
Wilmington	5	2
Winchester	3	3
Winthrop	2	1
Woburn	3-4	2-3
Wrentham	7	6

Bibliography

REFERENCES CITED

Abbreviations

AER, American Economic Review

AJE, American Journal of Economics

AJS, American Journal of Sociology

ASR, American Sociological Review

JAIP, Journal of the American Institute of Planners

JRSS-B, Journal of the Royal Statistical Society

RSA-P&P, Regional Science Association - Papers and Proceedings

SF, Social Forces

Books

Alonzo, W. Location and Land Use. Cambridge, Mass.: Harvard University Press, 1964.

Beshers, J. M. Population Processes in Social Systems. New York: Free Press, 1966.

______________ Urban Social Structure. New York: Free Press, 1962.

Blumen, I., M. Kogan, and P. J. McCarthy. Industrial Mobility of Labor as a Probability Process. Ithaca, N.Y.: Cornell U. Press, 1955.

Cooley, W. W., and P. R. Lohnes. Multivariate Procedures for the Behavioral Sciences. New York: Wiley, 1962.

Goldstein, Sidney. Patterns of Mobility: 1910-1950. Philadelphia: U. of Penn. Press, 1958.

Hanson, Hurwitz, and Madow. Sample Survey Methods and Theory, vol. 1. New York: Wiley, 1953.

Homans, G. Human Social Behavior: Its Elementary Forms. New York: Harcourt, Brace and World, 1961.

Hoover, E. M., and R. Vernon. Anatomy of a Metropolis. Garden City, NY: Doubleday, 1962.

Hoover, E. M. The Location of Economic Activity. New York: McGraw-Hill, 1948.

Kendall, M. G. A Course in Multivariate Analysis. New York: Hafner, 1957.

Laumann, E. O. Prestige and Association in an Urban Community. New York: Bobbs-Merrill, 1966.

Lawley, D. N., and A. E. Maxwell. Factor Analysis as a Statistical Method. London: Butterworths, 1963.

Loewenstein, Louis K. The Location of Residences and Work Places in Urban Areas. New York: Scarecrow Press, 1965.

Lowry, I. S. Migration and Metropolitan Growth. Chandler, 1966.

____________ A Model of Metropolis. Santa Monica: Rand, 1964.

Meyer, J. R., J. F. Kain, and M. Wohl. The Urban Transportation Problem. Cambridge, Mass.: Harvard U. Press, 1965.

Morgan, J. N., M. H. David, W. S. Cohen, and H. E. Brazer. New York: McGraw-Hill, 1962.

Quenouille, M. H. Introductory Statistics. London: Butterworth-Springer, 1950.

Rossi, Peter H. Why Families Move. Glencoe, Ill.: Free Press, 1955.

Schnore, L. F. The Urban Scene. New York: Free Press, 1965.

Tintner, G. Econometrics. New York: Wiley, 1965.

U.S. Bureau of the Census. US Census of Population: 1960, Detailed Characteristics of Massachusetts. Final Report PC (1)-23D Washington, D.C.: US Government Printing Office, 1962.

Vernon, R. Metropolis, 1985. Garden City, NY: Doubleday, 1963.

Articles

Alonzo, W. "A Theory of the Urban Land Market," RSA-P&P: 6(1960), 149-158.

Anderson, T. R., and L. L. Bean. "Shevky-Bell Social Areas: Confirmation of Results and a Reinterpretation," SF: 40 (Dec., 1961), 119-124.

_______________ and J. A. Egeland. "Spatial Aspects of Social Area Analysis, " ASR:26 (1961), 392-398.

Bell, W., and M. D. Boat. "Urban Neighborhoods and Informal Social Relations," AJS: 62 (1957), 391-398.

Berry, B. J. L. "Cities as Systems within Systems of Cities," in J. Friedmann and W. Alonzo, Regional Development and Planning, Cambridge, Mass.: MIT Press, 1964, Ch. 6, pp. 116-137; or, RSA:P&P, 13 (1964), 147-164.

Beshers, J. M., E. O. Laumann, and B. S. Bradshaw. "Ethnic Congregation, Segregation, Assimulation, and Stratification," SF:43 (1964), 482-489.

_______________ and E. N. Nishiura. "A Theory of Internal Migration Differentials," SF: 39 (1961), 214-218.

Brigham, E. F. "The Determinants of Residential Land Values," Land Econ.:41-4 (Nov., 1965), 325-334.

Carroll, J. D., Jr. "Some Aspects of the Home-Work Relationship of Industrial Workers," Land Econ.:25 (Nov., 1949), 414-422.

Duncan, B. "Factors in Work-Residence Separations," ASR:21 (1956), 48-56.

Duncan, O. D., and B. Duncan. "Residential Distribution and Occupational Stratification," AJS:60 (Mar., 1955), 493-503.

Feldman, A. S., and C. Tilly. "The Interaction of Social and Physical Space," ASR:25 (1960), 877-884.

Frieden, B. J. "Locational Preferences in the Urban Housing Market," JAIP:27 (1961), 316-324.

Gallaway, L. E. "Labor Mobility, Resource Allocation, and Structural Unemployment," AER:53 (Sum., 1963), 694-716.

Geary, R. C. "The Contiguity Ratio and Statistical Mapping," The Incorporated Statistician: 5, (1954), 115-145.

Goldstein, S. "Repeated Migratinn as a Factor in High Mobility Rates," ASR:19 (1954), 536-541.

___________ and K. Mayer. "Migration and the Journey to Work," SF: 42 (May, 1964), 472-481.

Greer, S. "Citizen Participation and Attitudes", in J. C. Bollens, ed., Exploring the Metropolitan Community. Berkeley and Los Angeles: U. of Calif. Press, 1961.

_________ "Urbanism Reconsidered: A Comparative Study of Local Areas in the Metropolis," ASR: 21, (1956), 19-25.

Horst, P. "Relations Among n Sets of Measures," Psychometrika: 26 (1961), 129-149.

Hotelling, H. "The Relation of the Newer Multivariate Statistical Methods to Factor Analysis," Brit. J. Stat. Psych.: 10 (1957), 69-79.

Kain, J. "Journey-to-Work as a Determinant of Residential Location," RSA-P&P: 9 (1962), 137-160.

Kish, L. "Differentiation in Metropolitan Areas," ASR: 19 (Aug., 1954), 388-398.

Kuh, E., and J. R. Meyer. "Correlation and Regression Estimates when the Data are Ratios," Econometrica: 23 (1955), 400-416.

Loewenstein, L. K. "Industry of Employment...," AJE: 24 (April, 1965), 157-162.

Lowry, I. S. "Location Parameters in the Pittsburgh Model," RSA-P&P: 11 (1963), 145-165.

Muth, R. F. "The Spatial Structure of the Housing Market," RSA-P&P: 7 (1961) 207-222.

Porter, A. "Approach to Migration through its Mechanism," Geografiska Annular: 38 (1956), 317-343.

Reeder, L. G. "Social Differentials in Mode of Travel, Time and Cost of the Journey to Work," ASR: 21 (1956), 56-63.

Ross, H. L. "Reasons for Moves to and from a Central City Area," SF: 40 (Mar., 1962), 261-263.

Schnore, L. R. "Socioeconomic Status of Cities and Suburbs," ASR: 28 (Feb., 1963), 76-85.

Shevky, E., and W. Bell. "Social Area Analysis," in G. A. Theodorson, ed., Studies in Human Ecology. New York: Harper & Row, 1961, pp. 226-235.

Suits, D. B. "Use of Dummy Variables in Regression Equations," J. Amer. Stat. Assoc.: 52 (1957), 548-551.

Tarver, J. D. "Occupational Migration Differentials," SF: 43 (Dec., 1964), 231-241.

Tintner, G. "Some Formal Relations in Multivariate Analysis," JRSS-B: 12 (1950), 95-101.

van Arsdol, M. D., S. Camilleri, and C. F. Schmidt. "An Application of the Shevky Area Indexes to a Model of Urban Society," SF: 37 (1958), 26-32.

Vyeki, E. S. "Residential Distribution and Stratification, 1950-60," AJS: 69 (Mar., 1964), 491-498.

Unpublished Material

Beshers, J. M. "The Construction of 'Social Area' Indices: An Evaluation of Procedures," Paper read at the ASA, Soc. Stat. Assoc., Dec., 1959.

Clake, W. L. "Intra-Metropolitan Migrants and Town Characteristics," unpublished MCP Thesis, MIT, 1967.

Henny, L. M. "Differential Educational Opportunities and Residential Choice..." unpublished MCP Thesis, MIT, 1966.

Leyland, G. P. "Migration to and within a Small Area," unpublished MCP Thesis, MIT, 1966.

Rust, E. C. "Intra-Metropolitan Migration: Six Boston Municipalities," unpublished MCP Thesis, MIT, 1963.