URBAN POPULATION CONCENTRATION AND OCCUPATIONAL STRUCTURE IN THE UNITED STATES: 1900-1950

By: Georges Sabagh, Maurice D. Van Arsdol, Jr., and Hamid Zahedi,
University of Southern California[1]

In recent years, an increasing number of studies have been focused on the interrelations between urbanization, modernization, and industrialization. In much of this research one methodological problem is the derivation of measures of urbanization which provide an adequate summary of the distribution of population by community size. This paper presents data to suggest that one logical and useful solution to this problem involves the derivation of measures of urban concentration from some mathematical model of concentration. An index based on the Lorenz curve model is presented and its relationship to other population distribution measures is documented. The research further describes the associations between a Lorenz curve index and measures of economic structure. The analysis is based on data for individual states of the United States from 1900 to 1950.

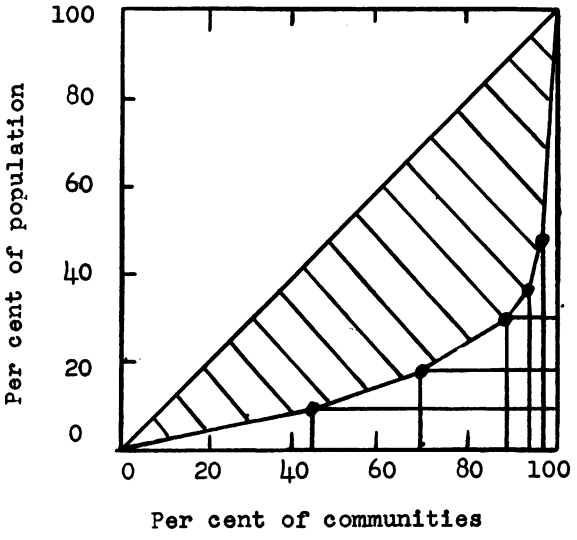
THE MEASUREMENT OF URBAN CONCENTRATION

Concentration indexes were first constructed for income size distributions. These indexes were originally developed by Pareto, Gini, Lorenz, and others, and have been applied to a wide variety of data, including size distributions of communities.[2] Two closely related expressions--Pareto's curve and the rank-size rule--have proven to be the most popular in urban research. However, analyses of the Pareto curve and rank-size rule have been more oriented toward describing empirical regularities and formulating size laws than toward deriving useful concentration indexes. Some investigators of the Pareto curve have realized that the slope of the curve can be used as an index of urban population concentration. To obtain such an index it is necessary to fit the curve to empirical data and then to measure the "goodness" and "closeness" of fit. If observed data depart from the Pareto model it is difficult to evaluate the empirical applicability of the slope as an index of urban population concentration. Concentration indexes that can be more simply derived include those based on the Lorenz curve. Few attempts have been made to measure urban population concentration on this basis, yet the method would appear to warrant further investigation.[3]

Structure of the Lorenz curve. The Lorenz curve describes the joint distributions of the cumulated proportions of communities of different sizes and of cumulated proportions of the populations of these communities. To the degree that the community and population size intervals are infinite, the Lorenz curve approaches a continuous curvilinear function. With a limited number of community and population size intervals it is discontinuous.[4] Figure 1 illustrates the Lorenz curve for community size and population for the United States as a whole in 1950. Coordinates on any of the

FIGURE 1

LORENZ CURVE FOR COMMUNITY SIZE AND POPULATION, UNITED STATES: 1950*
S=.35



*Based on "new" urban definition, and community size intervals of 2,500-4,999; 5,000-9,999; 10,000-24,999; 25,000-49,999; 50,000-99,999; and 100,000 and over.

points on the curve specify the proportions of communities and of populations below a given size level. The extent of urban concentration is proportional to the shaded area between the curve formed by the points and the diagonal. If the cumulated community and population proportions are equal, all of the observed points fall on the diagonal, all communities are of equal size, and minimum urban concentration is found. The observed points fall below the diagonal and a degree of urban concentration is obtained whenever the cumulated proportions of communities are greater than the cumulated proportions of populations. Complete concentration of population into a single city is found if the points of the coordinates delimit the total area of the triangle below the diagonal.

An index derived from the Lorenz curve. A number of indexes of urban population concentration can be obtained from the Lorenz model. The most common is one which delimits the area between the Lorenz curve and the diagonal of minimum population concentration, or equality of community size. This index, as used in the present analysis, is denoted as "S," and defined as follows:[5]

$$S = \tfrac{1}{2}\left[1 - (X_1Y_1) + (X_2-X_1)(Y_1+Y_2) + \ldots \ldots + (X_k - X_{k-1})(Y_{k-1} + Y_k)\right] = \tfrac{1}{2}\left[1 - \sum_{i-1}^{k} (X_i - X_{i-1})(Y_{i-1} + Y_i)\right]$$

where: X_i = the cumulated percentage of communities at point i; Y_i = the cumulated percentage of population at point i; and the range of S is .0 to .5. The S index value of .35 shown in Figure 1 indicates that the area between the Lorenz curve and the diagonal of minimum concentration is seventy per cent of the maximum possible area below the diagonal.

Computation of the S index for individual states. The S index was computed for each of the 48 states of the United States for the decennial census years 1900-1950. The six community size intervals used in these calculations ranged from 2,500 - 5,000 to 100,000 and over.[6] The choice of these intervals was guided by the availability of data and by the need for comparability with another analysis being prepared for various countries of the world.[7]

The S index values for individual states from 1900 to 1950 ranged from a minimum of .00 to a maximum of .43. Table 1 shows that the median values of the S index for states have increased slightly over the period 1900 to 1950 and reflect a trend toward urban population concentration documented by other investigators.

TABLE 1

MEDIAN STATE VALUES AND SEMI-INTERQUARTILE RANGE OF S, DAVIS, AND PER CENT URBAN INDEXES

UNITED STATES: 1900-1950

	S		Davis**		Per cent urban*	
Year	Mdn.	Q.	Mdn.	Q.	Mdn.	Q.
1950	.32	.08	.46	.10	.48	.09
1940	.32	.05	.47	.11	.42	.01
1930	.31	.06	.45	.11	.40	.13
1920	.30	.06	.50	.14	.36	.16
1910	.30	.06	.52	.15	.31	.15
1900	.27	.07	.59	.14	.28	.16

*Based on U.S. Bureau of the Census "old" definition.

**See footnote 9 for explanation of Davis index values.

Relations of the S index to other measures of urban population distribution. To evaluate the Lorenz curve as an independent measure of urban population distribution it was compared to two other indexes. One of them is the traditional census measure of the per cent of the population residing in incorporated communities with 2,500 or more inhabitants.[8] The other is an index comparable to one devised by Kingsley Davis to measure urban concentration.[9] This modified Davis index was obtained by taking an arithmetic average of the set of cumulated percentages of the urban population in communities of different

size, thus implicitly assigning an equal weight to each size interval. The census and the Davis indexes were calculated for each state from 1900 to 1950. The medians of these state values as well as the medians of the S indexes for states are given in Table 1. The intercorrelations of the state values of S with the Davis and percent urban indexes for 1900-1950 are shown in Table 2.

TABLE 2

ZERO ORDER CORRELATIONS OF STATE VALUES OF THE S, WITH THE DAVIS, AND PER CENT URBAN INDEXES
UNITED STATES: 1900-1950

Year	S-Davis	S-Per cent urban*
1950	-.75	.46
1940	-.51	.52
1930	-.40	.57
1920	-.40	.62
1910	-.32	.59
1900	-.21	.56

*Based on U.S. Bureau of the Census "old" definition.

The marked increase in the medians of the state values of the census index from 1900 to 1950 stands in definite contrast to the small change in S index values and a moderate increase in the values of the Davis index.

Table 2 indicates that the correlation between the S index and the per cent urban increased slightly between 1900 and 1920 and then decreased continuously until 1950. While the simple urban measure was once more closely related to the S index, the increasing amount of unexplained variance between them suggests that substitution of one measure for the other is not justified.

The marked increase in the correlation between the S and the Davis indexes from 1900 to 1950 may be explained by the fact that, with a fixed set of community-size intervals, the Davis index is more affected than the S index by the absence of communities in some of these intervals.[10] A consolidation of the upper two or three community-size intervals would have undoubtedly given a consistently higher correlation between the two indexes from 1900 to 1950. By 1950, when most states had communities in at least five of the six community-size intervals, the correlation given in Table 2 better reflects the relationship between the S and Davis indexes. The data for 1950 suggest that, in general, the S index correlates more closely with the Davis index than with the census index of per cent urban. Both the S index and the Davis index measure the extent to which the urban population of a state is concentrated in larger cities, whereas per cent urban more simply indicates the relative importance of all cities above 2,500 population. The main advantage of the S index over the Davis index is that it is based on a more general model of concentration.

URBAN CONCENTRATION AND OCCUPATIONAL STRUCTURE

The process of urbanization in Western nations has been linked closely with transformations in the basic nature of economic activity.[11] The initial shift from agriculture and home industry to factory-centered manufacturing has been accompanied by major currents in population redistribution. The next phase of economic modernization, characterized by the emergence of service and consumption oriented occupations, also witnessed significant population shifts. These economic changes can be portrayed, though somewhat crudely, by the changing importance of primary or predominantly agricultural occupations, secondary or manufacturing occupations, and service or tertiary occupations. One would expect, therefore, that there would be some association between measures of urbanization and indexes of occupational structure. In this respect, negative relationships are hypothesized between the level of urban population concentration and the proportion of the labor force in primary occupations, and positive relationships between urban concentration and the proportion of the labor force in secondary and service occupations.

Available published information

limits the scope of the description of the relationships between occupational structure and urban population concentration in the United States. By 1900, the first date for which measures of urban population concentration by states were derived, major changes in the growth and distribution of urban population had already taken place. Thus, the data to be presented are indicative of what may happen in a mature economy rather than in an economy in the initial phases of urbanization and industrialization.

Data on the occupational structure of each state of the United States in 1900, 1940, and 1950 were obtained from the University of Pennsylvania Study of Population Redistribution and Economic Growth.[12] The available occupational subclasses were combined to form primary, secondary, and service occupational classes. The proportions of gainful workers or persons in the labor force in each of the three occupational classes were computed for each state for 1900, 1940, and 1950.[13]

TABLE 3

ZERO ORDER AND MULTIPLE LINEAR CORRELATIONS BETWEEN STATE VALUES OF S INDEX AND PROPORTION OF THE LABOR FORCE IN PRIMARY, SECONDARY, AND SERVICE OCCUPATIONS
UNITED STATES: 1900, 1940, and 1950

	Years		
Variables	1900	1940	1950
S. primary	-.39	-.42	-.42
S. secondary	.23	.23	.21
S. service	.46	.34	.31
S. primary, secondary	.47	.44	.53
S. primary, service	.46	.43	.43
S. secondary, service	.49	.40	.42

Table 3 presents zero order and multiple linear correlations by states between the S index and the per cent of the labor force in primary, secondary, or service occupations for the years 1900, 1940, and 1950. While these correlations are low, they do form a pattern conforming to the hypotheses. For 1900, 1940, and 1950, the proportion of the labor force in primary occupations by states was found to be inversely related to the degree of urban concentration, without any changes over time. On the other hand, the correlations of secondary and tertiary occupations with urban concentration are positive for all years under consideration. An important historical change that is shown in Table 3 is the decreasing correlation of tertiary occupation and urban concentration over time. The same trend is noted in the multiple correlations when tertiary occupation is included as one of the independent variables.[14] In 1950 the highest multiple correlation was obtained for the combination of primary and secondary occupations.

The development of tertiary occupations may be partially a function of urban concentration which creates a need for service activities that are initially carried on within large cities. However, with the passage of time as the total society becomes more concentrated, and as techniques of transportation and communication are further developed, service activities may have become more rapidly dessiminated, allowing a deconcentration of both place of residence and place of work. This process may have been hastened by the development of white collar "dormitory" suburbs in recent years, which in the present analysis, would tend to increase the number of small communities and lower the value of the urban concentration index.

SUMMARY AND CONCLUSIONS

This paper has documented the importance of using measures of urban population concentration in the study of social and economic changes associated with urbanization. Indexes of urban population concentration derived from the Lorenz curve were described. It was shown that Lorenz curve index values tend toward independence from the traditional measure of per cent urban used by the United States Bureau of the Census. Possible uses of one of the Lorenz curve indexes in the study of changes in the economic structure of an urbanizing society were investigated. It was found that there is

a definite pattern in the relationship between urban concentration and the occupational structure characteristic of a maturing economy.

The findings of this study are tentative in nature. Further investigation is needed of the relationship of concentration indexes to other measures of urban population distribution and to more sensitive measures of economic development and structural changes in social organization. It would appear that research on urbanization needs to take into account more rigorous definitions of population distribution, and to focus on population redistribution as an important variable in the study of social change.

FOOTNOTES:

[1]This research was made possible, in part, by National Science Foundation Grant G 9452.

[2]Mary Jean Bowman, "A Graphical Analysis of Personal Income Distribution in the United States," American Economic Review, 35, 1945, pp. 607-628; and J. Aitchison and J.A.C. Brown, The Lognormal Distribution, Cambridge, Cambridge University Press, 1952. Chapter II. There is some evidence of the independent derivation of community size concentration indexes. For a discussion of early contributions in this area see: Otis Dudley Duncan, "The Measurement of Population Distributions," Population Studies, 11, 1957, pp. 28-30, and 39-42; United Nations, The Determinants and Consequences of Population Trends. New York, United Nations, 1953, pp. 175-176; and Eugenio D'Elia, Su alcuni metodi di calcolo della media aritmetica e della concentrazione nelle distribuzione multiple di frequenza. Roma, Studie e monografie della società Italiana di demografia e statistica, 3, 1949, pp. 10-18.

[3]E'Elia, op. cit., pp. 10-18, and "Methodologie statistique: la concentration," Bulletin de statistique (Belgium, Office central de statistique), 30, 1944, pp. 109-115. It should be noted that, by contrast, the Lorenz curve has been used extensively in the analysis of income distributions.

[4]It is possible to fit a curvilinear function to the observed data for the Lorenz curve, thus coming closer to a distribution that would have been observed with more detailed data. Such a procedure is not usually followed due to complications of curve fitting. Indexes derived from the actual data are approximations as they are based on a small sample of points.

[5]See: Duncan, op. cit., pp. 29-30; Aitchison and Brown, op. cit., pp. 112-113; and "Methodologie statistique: La concentration," op. cit., pp. 110-111. The notation given in the Belgian article was modified to conform with notation given by Duncan. Duncan's concentration ratio is identical with the S index, which is called "surface de concentration."

[6]See Figure 1 for a description of the size intervals used for the computations of the S index.

[7]Values of the S index may be affected by a change in the number of community size intervals used or by a modification in the lower or upper limits of the intervals selected. This problem is not given separate consideration in this paper.

[8]U.S. Bureau of the Census, U.S. Census of Population, 1950. Vol. II. Characteristics of the Population, Part I. United States Summary, U.S. Government Printing Office, Washington, D.C., 1952.

[9]Use of the Davis index is described in Kingsley Davis and Ana Casis, "Urbanization in Latin America," The Milbank Memorial Fund Quarterly, 24, 1946, pp. 186-207; Kingsley Davis, The Population of India and Pakistan, Princeton, Princeton University Press, 1951, p. 129; Jack P. Gibbs and Walter T. Martin, "Urbanization and Natural Resources: A Study in Organizational Ecology," American Sociological Review, 23, 1958, pp. 266-277.
Due to the manner in which the basic data were accumulated for the computation of the Davis index, Davis index values decrease with increasing urban population concentration and increase with decreasing urban population concentration.

[10]The number of states with no communities in at least three of the six intervals decreased from 15 in 1900, to 10 in 1910, 8 in 1920, 6 in 1930, 3 in 1940, and 1 in 1950.

[11]Colin Clark, The Conditions of Economic Progress, 3rd edition, London, Macmillan and Company, Ltd., 1957.

[12]Everett S. Lee, Ratner Miller, Carol P. Brainerd, and Richard A. Easterlin, Population Redistribution and Economic Growth, United States, 1870-1950. Vol. 1, Methodological Considerations and Reference Tables. Philadelphia, The American Philosophical Society, 1957, pp. 623-631.

[13]Ibid. Proportions are for gainful workers for 1900 and for the experienced labor force for 1940 and 1950. Primary, secondary, and tertiary gainful workers and persons in the experienced labor force are classified according to an interpretation of Colin Clark, op. cit. Agriculture, forestry, and fishing are classified as primary occupations; mining, construction, and manufacturing as secondary occupations; and transportation, trade, finance, services, and public administration as service.

[14]As the variance of a third occupational proportion is a function of the variance of the other two occupational proportions, only two occupational proportions were considered in each of the multiple regression equations.