Chapter 7

Industrial Location Analysis and Related Measures*

A. Introduction

In the previous chapters we have, more or less, assumed that an economic base, usually composed to a large extent of industry, exists in each region. We have touched on reasons for the existence of such industry. We have related population numbers, migration, Gross Regional Product and income, commodity and money flows, balance of payments, etc., to existing industrial bases and their characteristics. We have in addition considered cyclical effects and multipliers associated with change in the economic (industrial) base. However, nowhere have we probed with depth and with as much analysis as we can into the questions of what industries and how much of each can be expected to exist or develop in a region.

These questions of what industries and how much of each are basic to all forms of regional economic analysis. Fortunately, we can attack these questions with sharper tools than some of the questions we have already treated. We can go beyond mere description, which for example characterizes much work on regional income, population, migration, commodity and money flows, and balances of payments. We can get at more of the

* Sections B–E of this chapter were written with Eugene W. Schooler, and Appendices A and B with David F. Bramhall and Daniel O. Price.

relations governing decisions by business firms and government units. This consequence partly reflects the development of a considerable and rather sophisticated literature on location theory.[1] Although this literature is abstract and for the most part does not bear directly on the problems with which regional analysts and planners are concerned, it has led to at least one basic, general procedure which is exceedingly useful. This is the comparative cost technique, which also has roots in international trade theory. We shall discuss the comparative cost approach in the following section, an approach that casts considerable light on the "why" of systems of industrial locations. Then in succeeding sections we shall present materials on diverse types of coefficients and related concepts—the labor coefficient, the coefficients of localization and specialization, the localization curve, the index of diversification. These coefficients and related concepts are generally associated with location analysis. However, because they largely portray the "what" of systems of industrial locations, that is, are descriptive, they are not as useful as the comparative cost approach.

Finally, in Appendices A and B, we outline scaling and latent structure techniques and factor analysis, respectively. Scaling and latent structure techniques bear on community attitudes and other important subjective factors which must supplement a cost calculus for understanding systems of industrial location; in addition, these and related techniques promise to illumine several facets of regional behavior. Factor analysis, as discussed in Appendix B, pertains particularly to the proper delineation of regions within a system, a problem of major concern when coefficients of localization and related concepts are comprehensively employed to depict industrial location systems; too, factor analysis has potential applications in several forms of regional study.

B. Comparative Cost Approach

A comparative cost study typically proceeds for any given industry on the basis of an established or anticipated pattern of markets and a given geographic distribution of raw materials and other productive factors used in the industry. The objective of the study is to determine in what region or regions the industry could achieve the lowest total cost of producing and delivering its product to market. If the analyst is concerned with the industrial growth prospects of a particular region, a series of such comparative cost studies is rather essential.

[1] Among other literature on location theory, the reader is referred to J. H. von Thünen [65]; W. Launhardt [46, 47]; A. Weber [68]; T. Palander [53]; A. Lösch [52]; W. H. Dean, Jr. [11]; E. M. Hoover [34]; E. S. Dunn [16]; M. L. Greenhut [24]; W. Isard [38]. There is an elementary statement on location analysis in E. M. Hoover [33].

Frequently the initial justification for one or more comparative cost studies arises because of changes in general technology, or in the technology of a particular industry, or in the production of an individual raw material or intermediate good. For example, general improvement in a region's internal and external transport situation—the completion of a system of superhighways, the erection of more modern and efficient railroad terminal facilities, the construction of deep-water river channels, etc.—can have a significant effect on the relative advantages of the given region for industrial location. If the region is lightly populated with a plentiful variety and supply of natural resources, the transport improvement could tip the scales of regional advantage in its favor. Such improvement might give the region an advantage over locations using inferior or high-cost raw materials which had nevertheless *been* best because of nearness to markets. On the other hand, if the region itself is a densely populated market area, the transport improvement might drastically cut costs of assembling raw materials there. Thus, in certain heavy raw-material-using activities, such improvement would allow a shift in advantage from raw material regions to the region under consideration.[2] In any case, because the transport improvement affects all industry in general, the regional analyst should pursue comparative cost studies for a number of industries.

Another situation that could be usefully analyzed by means of one or more individual-industry comparative cost studies might arise because of a change in market conditions. For example, as a region grows and develops, its population expands. Its local market becomes capable of absorbing the outputs of economic-size plants in a growing number of industries. Comparative cost studies can indicate for which of these industries local production can be justified.

Still another situation might be associated with a changing raw material supply pattern, for example, from the gradual depletion of a locationally dominant ore source. Or it might be associated with a prospective new industry, such as nucleonics or electronics, or with a new productive process, such as irradiation or continuous casting. In each of these situations a regional comparative cost study, taking these changes into account, can help to indicate whether there is a basis for a relocation or new growth of industry and, if so, the nature of the location patterns to be expected.

With these comments illustrating some of the many possible situations in which individual-industry comparative cost studies are useful, we turn

[2] As an instance, the relatively recent construction of large-diameter, long-distance natural gas pipelines has provided a practical possibility for the production of ammonia and other natural-gas-based petrochemicals in regions far from the natural gas fields.

to procedures for conducting such studies. The most direct way to pursue a comparative cost study for an industry would be to secure enough information to calculate the total production costs the industry would incur in each of the regions to be compared. The region or regions with the lowest production costs (including transport cost) would be the most desirable location, in an economic sense. Since the difference in total cost from region to region is the important magnitude, it becomes clear with further reflection that the regional comparative cost study need consider only the production and transport cost elements which actually differ from region to region. The components of production and transport cost that do not vary regionally in amount may be ignored; they give rise to no regional advantage or disadvantage. In practice, this consideration of cost differentials only can lead to considerable saving of research time, since many items of production cost for most industries do not exhibit systematic or significant regional variation.[3]

It should also be observed that in considering an element of production cost which does vary regionally, it is often possible to estimate the amount of its difference between regions without knowing its absolute regional levels. For example, take two similar plants, one in New York City, the other in a coal town near Pittsburgh. Each consumes ten tons of coal a day, the cost of coal in New York City tending to exceed the cost of coal in the coal town by the cost of transporting coal from the coal town to New York City. If it is known that the transport rate is $3 per ton, the daily coal cost of the New York plant would exceed by $30 the daily coal cost of the other plant. However, this method of computing a regional cost differential can be used only if the relevant productive factor input is the same in each region, both in type and quantity. Thus, if the New York plant used only eight tons of coal a day as compared to ten in the other plant, the analyst would have to know the price or cost of coal in at least one of the sites as well as its transport cost in order to compute the daily coal cost differential. Also, if one of the plants adopted a productive process that used electricity rather than coal, the analyst would need to know the absolute cost of both electricity and coal in order to calculate the energy (fuel and power) cost differential between the two plants.

1. PETROCHEMICAL LOCATION

To illustrate the comparative cost technique, which in essence involves

[3] Also of practical importance is the fact that industrial companies are often willing to furnish information regarding a few individual production cost items but refuse to divulge a complete itemized summary of unit production costs.

a systematic listing of regional cost differentials,[4] we summarize a recent analysis of factors affecting the future location pattern of the natural gas-based petrochemical industry.[5]

The first task of the analysis was to determine which components of petrochemical production cost could be expected to vary regionally. Generally, these components are fuel and raw material gas, steam, electric power, labor, and transportation. Additionally, major cost differentials result if feasible sizes of productive units differ regionally. (Large units achieve "economies of scale" which are denied to small units.)

The second step of the analysis was to choose and define the regions to be compared. Consideration of various geographic and technological factors led to the conclusion that feasible locations for the production of petrochemicals are limited to (1) sites in the Gulf Coast and adjoining interior area (in which are concentrated most of the country's reserves of natural gas) and (2) several sites within the country's major petrochemical market area generally extending from the industrial Northeast through the Great Lakes region.[6] This over-all market area was divided into several smaller market regions, each constituting the natural hinterland or distribution area of a city strategically located with respect to interregional transport connections. The relevant regional cost comparisons were thus between (1) a site in the Gulf Southwest raw material region and (2) the focal or distributional point site in each of the several market regions.

The following tables summarize the results of the calculation of regional cost differentials in the production of a typical petrochemical, *ethylene glycol* (the basic component of permanent-type antifreeze).[7] Table 1 lists the raw material, utilities, and labor inputs which may lead to regional cost differentials. The cost comparisons are between a Mississippi River location in the raw materials region near Monroe, Louisiana, and a market region location at Cincinnati, Ohio.

It is demonstrated in the study that regional differentials in the cost of ethane (a raw material gas contained in natural gas) may be viewed as approximately equal to regional differentials in the cost of the equivalent volume of natural gas. Similarly, since the gas used for fuel is natural gas, regional differentials in the cost of process fuel gas are regional differentials in the cost of natural gas. Further, differences in steam costs

[4] For the formal, theoretical approach see W. Isard [38]. A brief, step-by-step exposition of a modern Weberian comparative cost approach as applied in industrial complex analysis is found in Chapter 9.

[5] W. Isard and E. W. Schooler [42].

[6] Other possible sites for the production of certain petrochemicals are within the Pacific Southwest region.

[7] These tables are based on figures appearing in similar tables in W. Isard and E. W. Schooler [42], pp. 19, 22–24.

depend mainly on fuel cost differences. This means that regional steam cost differentials can be expressed as equal to the regional differentials in the cost of the required fuel gas (natural gas). Thus raw material gas, fuel gas, and steam cost differentials can be combined into a net cost differential on the equivalent volume of natural gas. This net cost differential will in the long-run tend to equal the interregional pipeline transport cost on that volume of natural gas.

A market site location incurs transport inputs (costs) on raw material gas, process fuel gas, and fuel gas for steam; practically speaking, it avoids transport inputs (costs) on finished product. A raw material site location avoids transport inputs (costs) on raw material gas, process fuel gas, and fuel gas for steam; it incurs transport inputs (costs) on finished product.

TABLE 1. PRODUCTION OF ETHYLENE GLYCOL FROM ETHANE
(Via oxidation process)

Selected Inputs	Requirements per 100 Pounds of Ethylene Glycol
Ethane	108 lb.
Utilities	
Fuel gas	377 cu. ft.
Steam	1248 lb.
Electric power	10 kw-hr.
Labor	0.19 man-hours

Because of their different requirements of transport inputs, a comparison must be made in order to calculate the transport cost differential which may exist between these two locations. This comparison is presented in Table 2.[8] Monroe has a net transport cost advantage of 13 cents if large-volume barge shipment is possible. Cincinnati has a net transport cost advantage of 60 cents if the finished product must move by rail.

Another major cost differential in petrochemical production is associated with differences in plant size. Table 3 presents the results of a calculation of such scale economies. These economies may amount to as much as $4.00 per 100 pounds of ethylene glycol.

[8] Transport cost on fuel and raw material gas was calculated on the basis of the ethane, steam, and fuel gas requirements shown in Table 2 converted to their natural gas equivalents as follows:

1 pound ethane—12.7 cubic feet
1 pound steam—1.5 cubic feet
1 cubic foot fuel gas—1 cubic foot

The interregional pipeline transport rate on natural gas was taken to be 1.3¢ per thousand cubic feet per hundred miles.

TABLE 2. TRANSPORT COST DIFFERENTIALS PER 100 POUNDS

Case A. Shipment of Product by Barge

Location	Transport Cost on: Equivalent Natural Gas	Transport Cost on: Finished Product	Total Transport Cost	Net Advantage of Monroe
Monroe	0	16 ¢	16¢	13¢
Cincinnati	29¢	0	29¢	

Case B. Shipment of Product by Rail

Location	Transport Cost on: Equivalent Natural Gas	Transport Cost on: Finished Product	Total Transport Cost	Net Advantage of Cincinnati
Monroe	0	89¢	89¢	
Cincinnati	29¢	0	29¢	60¢

Other possible cost differentials which may be significant are those in direct labor and electric power. With respect to these two items, the estimated maximum possible cost differentials between any two regions in the United States are presented in Table 4. For the two regions actually being compared, analysis indicates that differentials in power cost and labor costs will be so small as to be virtually insignificant. This leaves transport cost differentials and economies of scale as the important factors in this comparative cost study. How can we interpret the materials on these factors?

TABLE 3. ECONOMIES OF SCALE PER 100 POUNDS ETHYLENE GLYCOL ASSOCIATED WITH DIFFERENT SIZES OF ETHYLENE-ETHYLENE GLYCOL PRODUCTION UNITS

Scale economies of medium plant over small plant	$2.45
Scale economies of large plant over medium plant	$1.53
Scale economies of large plant over small plant	$3.98

Initially it should be noted that feasible plant size at a natural gas site location will generally be as large or larger than the feasible plant size at a market site because a natural gas site can serve various markets whereas a market site can serve efficiently only its own market area. Therefore, any regional cost differential due to economies of scale will always tend to favor a natural gas site location.

If the market demand were large enough to justify the shipment of ethylene glycol to market by river barge, it would be large enough to absorb the output of at least one large optimum-size market site plant. Hence, the ethylene glycol production serving this market would come from a large-size plant, whether the plant was at the market site or the natural gas site. This would mean little or no regional cost differential due to economies of scale. The net transport cost differential would constitute the entire regional cost differential. In the case depicted by Table 2 it would amount to 13 cents per hundred pounds of ethylene glycol, favoring a natural gas site location.

If the market demand were small, that is, if it were insufficient to justify shipment of product by barge, a potential natural gas site plant would have to ship by railroad tank car. Under such conditions, a potential market site plant would possess a decided transport cost advantage, but at the same time the smallness of the market demand would indicate that the market could absorb the output of only a small-size market site plant. There would exist regional cost differentials stemming from economies of scale, since the potential natural gas site plant could be of large or at least medium size because of its possibility of serving multiple markets. A comparison of the figures in Table 4 and Table 3 shows that the scale advantage of even a medium-size plant at Monroe (which advantage is $2.45) will much more than offset the transport advantage of a small plant at Cincinnati (which advantage is $0.60). The resulting net regional cost differential in favor of the Monroe location is $1.85 per hundred pounds of ethylene glycol.

TABLE 4. MAXIMUM LABOR AND POWER COST DIFFERENTIALS PER 100 POUNDS

Maximum labor cost differential	12¢
Maximum power cost differential	6¢

The general conclusion supported by this regional comparative cost study is that a natural gas raw materials region site on the Mississippi River near Monroe, Louisiana, is the better location for producing ethylene glycol for the Cincinnati market area.

It is evident from these materials and other extensive sets of data that in a regional comparative cost study of the production of petrochemicals from natural gas, regional differences in total transport costs and possible regional differences in feasible plant sizes are the only major considerations. In one sense, regional differences in plant sizes, owing to limited individual market demand, are of overriding importance. When these differences exist the resulting scale economies of the large plants in the raw-materials region will in virtually every case completely overshadow any other individual or combined regional cost differential. However, for all the principal volume petrochemicals there are individual market regions, each of which encompasses a demand sufficient to absorb the output of at least one optimum-size plant. In such cases there can be no appreciable regional cost differentials from economies of scale, and total transport cost differentials become the only significant regional location factor.[9]

2. IRON AND STEEL LOCATION

An essentially similar picture is presented by analysis of the iron and steel industry. A study of the feasibility of a New England location for an integrated iron and steel works may be used as an illustration.[10] Considering first the obviously important factor of transport cost differentials, the study presents tables showing total transport costs on major raw materials and finished steel products incurred by various actual and hypothetical production locations in serving various New England market centers. Table 5 is the transport cost table applying to the Boston market.[11]

The figures show that for serving the Boston market either of the two New England locations considered, Fall River or New London, has a net transport cost advantage over other locations. For example, Pittsburgh incurs a total transport cost of $22.31, and Fall River and New London incur costs of $13.91 and $15.90, respectively (when Labrador ore is smelted). Similar results appear in the tabulations for other New England market centers, with the exception that for those in southern and western New England the transport cost advantages of Fall River and New London over Trenton are sharply reduced and under some conditions disappear.

TABLE 5. TRANSPORTATION COSTS ON ORE AND COAL REQUIRED PER NET TON OF STEEL AND ON FINISHED PRODUCTS FOR SELECTED ACTUAL AND HYPOTHETICAL PRODUCING LOCATIONS SERVING BOSTON

Location		Transportation Costs on:			
		Ore	Coal	Finished Product	Total
Fall River	Labrador ore	$4.56	$6.01	$4.60	$15.17
	Venezuela ore	3.68	5.63	4.60	13.91
New London	Labrador ore	4.56	5.79	6.80	17.15
	Venezuela ore	3.68	5.42	6.80	15.90
Pittsburgh		5.55	1.56	15.20	22.31
Cleveland		3.16	3.85	15.20	22.21
Sparrows Point		3.68	4.26	12.40	20.34
Buffalo		3.16	4.27	12.60	20.03
Bethlehem		5.56	5.06	10.60	21.22
Trenton		3.68	4.65	10.40	18.73

The study proceeds by analyzing other production costs. Most of these are shown to be subject to no significant regional variation. Labor costs in the highly unionized iron and steel industry are effectively equalized among regions. It is stated that taxes may vary significantly from state to state and from locality to locality, but that there is no basis for estimating

[9] The theoretical schema which formalizes the methodology of the comparative cost study is substitution analysis as developed in the first volume on *Location and Space-Economy*. Essentially this analysis considers alternative locations in terms of substitution between transport inputs, between diverse outlays, between diverse revenues, between outlays and revenues, and between combinations of these substitutions. The best location, in an economic sense, is one where no move elsewhere could result in further favorable substitution, that is, in reduction in total production and delivery cost. Each of the regional cost differentials of a comparative cost study measures the effect of either a single or combined substitution involved in the decision to locate in one region rather than another. In the petrochemical industry case, the only relevant regional substitution possibilities were between transport inputs (if market demand is large and concentrated), or between transport inputs and between outlays on transport inputs and outlays on production in general (if market demand is small and scattered). Because of various technological and economic reasons, the analysis of the petrochemical industry was limited to a comparison of location at a raw material site versus location at a market site. The formal substitution analysis encompasses the general case of multiple location possibilities, including intermediate sites, and variable factor proportions.

A graphic presentation of the substitution analysis can be accomplished by means of the isodapane technique. Isodapanes are essentially contour lines showing loci of points of equal production and delivery cost, equal additional transport cost, equal percentages of some base amount, etc. See T. Palander [53] and E. M. Hoover [34].

[10] W. Isard and J. Cumberland [41].

[11] This is Table II in W. Isard and J. Cumberland [41], p. 249. A detailed explanation of the sources of the estimates and the assumptions under which they were derived appears as a note to Table I, p. 248.

the amount and direction of such variation. It is also pointed out that a New England location would enjoy an initial advantage because of lower prices on scrap iron and steel, but that this would tend to disappear when a New England steel mill became a major scrap user. And so forth.

Finally, account is taken of the influence exerted by the size of the New England market demand. In order to achieve economies of scale and juxtaposition, each productive unit of an integrated steel works must be of at least minimum economic size. The demand and capacity estimates used in the study indicate that, although the total New England steel demand would well exceed the total tonnage output of an efficient integrated steel works, it is uncertain that the demand for each of the specialized components of total output would be sufficient to absorb the output of an economic-size unit. Because of the uncertainty of market demand magnitudes and in view of other intangibles, the study concludes that the net regional advantage of a New England location is not proved, even though it does enjoy a significant transport cost advantage.

3. ALUMINUM AND OTHER INDUSTRY LOCATION

The regional location patterns of the petrochemicals industry and the iron and steel industry are influenced primarily by transport costs, given the existence of large-scale individual market demands. They are essentially transport-oriented industries. A somewhat different situation exists with respect to the aluminum industry. From the standpoint of transport costs alone, the best locations within the United States for serving several of its major industrial aluminum markets are generally market locations.[12] Yet there is little aluminum production capacity at major market centers. Clearly there is some locational influence at work which is stronger than that exerted by regional transport cost differentials. It proves to be the influence of regional differences in the cost of electric power. For illustration, a location in the New York City area, a major market center, may be compared with a location in the Pacific Northwest, a region possessing major aluminum reduction capacity. Table 6 compares regional transport costs alone.[13]

Now consider the influence of the power cost differential. The production of 1 pound of pig aluminum requires approximately 9 kilowatt-hours of electric power. This means that a difference of 1.91 mills (which equals 1.714 cents divided by 9) per kilowatt-hour in the cost of electric power would be enough to offset completely the net transport cost advantage of the New York location. Actually, power rates in the Pacific Northwest

[12] W. Isard and V. H. Whitney [43].

[13] These transport cost figures are taken from W. Isard and V. H. Whitney [43], Tables XXII, XXIII, pp. 125, 127.

range from 2.5 to 3.5 mills per kilowatt-hour whereas in the New York area rates are approximately 8 mills per kilowatt-hour.[14] Thus, any advantage that New York possesses on transport cost account is clearly overshadowed by its disadvantage on power cost account.

Other regional cost differentials in the production of aluminum are relatively slight. As a result, the aluminum industry can be expected to be located and grow in a cheap-power region. It is a power-oriented industry.

With respect to other industries, analysis may show that some other production cost component gives rise to a major regional cost differential.[15] For example, the dominating locational influence in the textile industry is exerted by regional differentials in labor costs. But whatever the relative

TABLE 6. REGIONAL TRANSPORTATION COSTS: ALUMINUM PRODUCTION FOR NEW YORK MARKET

Item	Location at: New York	Pacific Northwest
Transport costs per pound pig aluminum		
a. On raw materials	0.548¢	1.312¢
b. On pig aluminum	0.000	0.950
Total	0.548¢	2.262¢
Net transport cost advantage of New York: 1.714¢		

importance of the various types of regional cost differentials, the general approach to the regional comparative cost study is the same. The analyst must identify the components of production cost which vary regionally and then estimate the amount of each resulting cost differential. Finally, calculation of net differentials will identify the region or regions in which the industry would enjoy minimum production costs, given the set of simplifying assumptions which must underlie such a study.

[14] These rates are characteristic of 1955 conditions.

[15] In addition to the references cited in connection with the illustrations in the text, particular industry studies which illustrate the comparative cost approach are J. V. Krutilla [44], J. R. Lindsay [50, 51], J. Airov [2, 3], E. W. Schooler [58], W. Isard and W. M. Capron [40] and J. Cumberland [9].

Summaries of comparative cost studies and discussion of comparative cost factors in various industries are found in, among others, S. H. Schurr and J. Marschak [59], W. Isard [39], and T. R. Smith [62].

4. SOME LIMITATIONS

In this manner a researcher may systematically pursue comparative cost analysis for each industry considered relevant for a region. In doing so, it may be said that in at least one sense he is effectively studying the internal industrial structure of a region. Further, if he extends his analysis for each industry to embrace all regions in a system, it may be said that he is effectively studying the internal structure of the system (and as a necessary consequence the internal structure of each region).

Upon reflection, however, it is seen that this statement of possible achievement is misleading. It is to be recalled that in a comparative cost study for a given industry with reference to a specific region, both the price-cost structure and the magnitude of the market existing in each region are assumed given. Where the given industry is small and has little influence on income, demand, prices, and costs in any region, these assumptions may be justifiable. But such assumptions are clearly not warranted when the geographic pattern of the industry does have a marked influence on income, demand, prices, and costs in one or more regions. And certainly these assumptions are untenable when the researcher purports to analyze locationally each industry relevant for a region, since the estimated income and markets of the region and much of its price-cost structure is largely contingent upon the amount of industry to be located in a region.

These remarks point up the need to supplement comparative cost analysis, *when it is pursued on a systematic basis*, with other techniques which are aimed at uncovering interrelations of industry and the mutual dependence of their markets. These techniques, such as interregional input-output, industrial complex analysis, and interregional linear programming will be discussed in subsequent chapters. When coupled with comparative cost analysis, it will be seen that they can provide greater insight into the interindustry structure and other interrelations of a system of regions.

Despite the promise of comparative cost analysis when combined with the techniques mentioned, results which may be obtained must be qualified with respect to at least another major factor. It frequently happens that a comparative cost study points up a particular area as an ideal location for a given industry. Yet because of the resistance of the business units, social groups, and household residents of the area—which resistance may be formal (e.g. zoning restrictions), informal, or both—the industry does not locate in the area. More broadly speaking, from a cost standpoint a region may be ripe for industrial development. Yet because of native attitudes, cultural patterns and institutions, and other noneconomic factors, attempts at industrial development are aborted. Ideally, regional analysis should incorporate the play of such noneconomic factors which are largely nonquantitative in character. Unfortunately, only little can be done in this direction at the present time. What can be achieved, drawing upon scaling, latent structure, and similar techniques developed by psychologists and sociologists, is sketched in Appendix A to this chapter.

C. The Labor and Similar Coefficients

It was indicated in the preceding discussion that a series of regional comparative cost studies is an effective analytical tool. The researcher can use this tool to appraise the locational attractiveness of a region which possesses an abundance or cheap source of some particular mineral, commodity or service input, or other factor or market advantage. In a "cheap labor" region, for example, each individual industry comparative cost study would not only indicate in quantitative terms the pulling power of cheap labor for that industry but would also indicate whether or not the region's cheap labor advantage is sufficient to outweigh any locational disadvantages it might suffer compared to other regions. The analyst could, from an examination of the series of comparative cost studies, compare the net locational effect, industry by industry, of the region's cheap labor.

It is quite possible, however, that the analyst would have neither the time nor the resources to carry out a thorough series of regional comparative cost studies. He might desire to short-cut the extensive computations by using various coefficients. He might consider, as a first step, an industry-by-industry calculation of average labor costs per dollar of output.[16] The larger this value, the larger would be the absolute labor cost differential per dollar of output associated with a given regional wage rate differential. Thus, it would appear that industries most likely to locate in the cheap-labor region are those with the highest average labor cost per dollar of output. However, a few moments' reflection can usually bring to mind a number of instances in which industries with comparatively high labor costs per unit or per dollar output have not established production in possible cheap-labor areas but have continued to expand operations in regions of relatively dear labor. At the same time it may be quite possible to point out cases of industries which have lower labor costs per dollar output but which have actually been attracted by the cheap-labor regions. The logical explanation for such situations is that there are costs other than labor costs which vary regionally, and that these other cost variations

[16] Variants would be labor cost per unit of output and labor cost as a percentage of total unit cost.

are of different magnitudes for different industries and thus exercise varying degrees of locational influence. The problem in assessing the relative strength of a cheap-labor region's attraction for various industries lies in devising some method to take account of differences not only in labor costs among industries but also in other cost items.

One quite universal element of cost which varies significantly among industries and which, for any given industry, is generally subject to persistent regional variation is that of transport cost. To indicate the relative attractiveness of a cheap-labor location for different industries, with due regard to interindustry transport cost differences, Weber developed his "labor coefficient." It is the ratio of the labor cost per unit of product (at existing locations) to the "locational weight" of that unit. The locational weight is the sum of the required weights of localized raw materials plus product.[17] Other things being equal, the higher an industry's labor coefficient, the more likely it is that the labor cost savings it could achieve in a cheap-labor area will exceed the additional transport costs incurred by not locating at a minimum transport cost site. Generally speaking, the locational attraction of cheap-labor areas is greater for industries with high labor coefficients than for those with low.

Although the method of ranking industries by their labor coefficients affords a useful priority list of industries from the standpoint of their attraction to cheap-labor areas in general, it has definite limitations. First, more information than the labor coefficient is required to determine whether a given industry should actually be established in a given region, even if labor costs and transport costs are the only significant locational variables. The numerator of the labor coefficient must be multiplied by the "percentage of compression" of wage rates achieved by the cheap-labor region relative to the rate used in computing the labor coefficient.[18] This yields the labor cost saving per unit of product. Then the denominator (locational weight per unit product) must be multiplied by the transport rate and the net additional distance involved in location away from a minimum transport site. This yields the additional transport cost incurred

[17] Ubiquitous raw materials (ubiquities) are not included in locational weight, since in their unprocessed state they never need be transported.

All weights are expressed as "ideal weights," that is, actual weight adjusted so as to have the effect of equalizing transport rates on all materials and product. (E.g., a ton of a commodity which incurs a rate twice as great as a standard commodity is considered to have an ideal weight of two tons.) In the use of this coefficient, transport costs are generally assumed to be proportional to distance.

For supplementary discussion, see *Location and Space-Economy*, pp. 126–142.

[18] The percentage of compression thus represents the per cent by which the wage rate at a cheap-labor location is lower than the wage rate at existing locations (after adjustment to an equivalent efficiency basis).

by locating in the cheap-labor region. Only if the labor cost saving is greater in amount than the additional transport cost would it be to the industry's advantage to locate in the cheap-labor region, *ceteris paribus*.

Second, consider the relative attraction held by a specific cheap-labor area for two industries with identical labor coefficients but with geographically different minimum transport cost production sites. Except for special cases, the relative pull of the cheap-labor area would be different for the two industries; it is quite possible that location there would represent a net advantage for one but a net disadvantage for the other, depending on the net additional distance involved in location away from the respective minimum transport cost sites.[19]

It is even possible that the distance factor could be so different for two industries with reference to a specific cheap-labor site that one could have a relatively low labor coefficient and yet be attracted to the site, and the other could have a relatively high labor coefficient and yet tend to locate away from the site. It becomes evident that even when labor costs and transport costs are the only significant locational influences, a ranking of industries by their labor coefficients is a valid indicator of the relative degree to which they are attracted to a specific cheap-labor site only when the net additional distance involved in a location away from the minimum transport cost site is the same for each industry.[20] Such a ranking can be definitely misleading if the distances involved vary among industries, as they commonly do. The ranking can be still more misleading if we consider too the possibility of transport savings at a cheap-labor site from the use of substitute sources of raw materials, since this possibility may vary greatly from industry to industry and from region to region.[21] The relative degree of the cheap-labor site's attraction for different industries can be assessed under such conditions only by individual calculation for each industry of the labor cost saving per unit compared with the additional transport cost per unit.

[19] For example, a cheap-labor area in, say, South Carolina might attract an industry utilizing lower Mississippi Valley raw materials and serving a Middle Atlantic industrial market; yet for an industry having an identical labor coefficient but using raw materials from the Central Plains states and serving a Chicago-Detroit market, the labor cost savings attainable at the South Carolina location could very well be entirely inadequate to offset the additional transport cost incurred on the shipment of raw materials and product.

[20] Also, in Weberian terminology, the location figures must be identical (or, practically speaking, approximately equal).

[21] Additionally, the possibility of cost savings from substituting cheap labor for other factor inputs may exist, and to a different extent from industry to industry. Such cost savings must be allowed for in the final cost comparisons.

The reader will perceive that this sort of calculation is actually an industry comparative cost study in an abridged form. That is, cost differentials are calculated between two regions, the region of the minimum transport cost site and the region of the cheap-labor site, with respect to only two cost elements, transportation and labor. This illustrates the basic limitation of the labor coefficient and of any comparative cost calculation derived from it. At the end, the analyst finds himself with a set of incomplete regional comparative cost studies. A better method would be to proceed from the outset with a more complete and systematic set of industry-by-industry comparative cost studies (embracing, if possible, recognition of community attitudes and other similar subjective factors discussed in Appendix A to this Chapter).

Of what use then is the labor coefficient? It has already been noted that a ranking of industries by their labor coefficients furnishes an indication of the extent of their attraction to cheap-labor regions *generally*. To the analyst concerned with the growth prospects of a specific cheap-labor region, such a ranking is an aid in deciding for which industries to carry out comparative cost studies. Certainly those with high labor coefficients would be included at the start. The ones with lower coefficients should not be summarily rejected, but the analyst would realize that their growth in his region, in all likelihood, depends on other factors in addition to possible labor cost savings.

Although the labor coefficient has been discussed to illustrate a device useful to a limited extent in the analysis of a region possessing a particular resource, similar coefficients can be developed in connection with other specific regional resource advantages. Thus, a power coefficient, a fuel coefficient, a steam coefficient, among others, could be developed to indicate the relative extent to which various industries are attracted to regions of cheap power, fuel, steam, etc. These coefficients would, as in the case of the labor coefficient, have as a numerator the average cost of the specific resource input per unit output of product, and as a denominator the locational weight associated with the unit output of product.[22] Their use would likewise be principally as general indicators rather than as specific measuring devices.

If a region possessed more than one resource advantage, for example, cheap labor *and* cheap power, a combined coefficient might perhaps prove useful in certain situations. The numerator in this case would consist of the combined labor and power cost per unit output, and the denominator would again be the locational weight of the unit output of product. However, the two elements of the numerator would have to be weighted, the weights for any region being the respective percentage compressions achievable in the region for these elements. Since the weights would differ from region to region, the usefulness of a combined coefficient is greatly curtailed.

[22] In the computation of these as well as labor coefficients, it is sometimes possible and desirable to take account, in the *denominator*, of the effects of using such resources as power, fuel, and steam. Often regional differences in power and steam costs are attributable almost wholly to regional differences in the cost of the fuel required; these fuel differences, in turn, may simply reflect the cost of transporting fuel such as coal or gas from one region to another. If the process power, steam, and fuel requirements are expressed in terms of their required fuel equivalents, the latter can then be considered as part of the locational weight. For a more complete discussion of this point, see W. Isard and E. W. Schooler [42], pp. 15–16.

If regional differences in power or steam costs are wholly due to differences in transport cost between regions of required fuel, a set of power coefficients or steam coefficients indicate, in effect, the relative importance, among the different industries, of transport inputs associated with power and steam requirements, compared with other transport inputs, in establishing the minimum transport cost site.

D. Coefficient of Localization, Localization Curves and Ratios, and Related Concepts

The coefficients discussed so far are primarily applicable to the analysis of a region with an abundant and cheap endowment of one or more particular resource. However, the regional analyst may be concerned not so much with finding which industries can best use an abundant resource as with finding industries to diversify the economic base of the community. Or he may be concerned with possible lines of development in a region committed to a specific policy of small industries or small plants or both. Or he may be concerned with the change over time of the spatial pattern of population and total employment, or with the change over time in the degree to which one or more industries are material- or market-oriented.

1. Preliminary organization of data

To help deal with such concerns and problems and many similar ones, a number of coefficients, ratios, and indexes have been developed. Many of these pertain to the same sets of data. It therefore is desirable to discuss them in a rather systematic manner. To facilitate this discussion, we sketch the outlines of a table containing certain basic data.

Table 7 relates to 1954 manufacturing employment by industry for the United States viewed as a system of regions. (The problem of selecting appropriate sets of regions and industries will be discussed later and in Appendix B.) Each column refers to a particular state (region) of the United States; the total manufacturing employment in each state (region)

TABLE 7. LOCATION QUOTIENTS AND EQUIVALENT EMPLOYMENT PERCENTAGE RATIOS BY REGION AND INDUSTRY, 1954

	Employment: United States	Maine	New Hampshire	. . .	Indiana	Michigan	. . .	California
Employment (total)	16,125,550	104,507	77,332	. . .	587,782	1,056,564	. . .	1,052,785
Meat products	311,336	$\frac{0.36}{0.65}=0.55=\frac{1.06}{1.93}$	$\frac{0.15}{0.48}=0.31=\frac{0.59}{1.93}$	. . .	$\frac{3.54}{3.65}=0.97=\frac{1.88}{1.93}$	$\frac{2.00}{6.55}=0.31=\frac{0.59}{1.93}$	. . .	$\frac{5.76}{6.53}=0.88=\frac{1.70}{1.93}$
Dairy products	283,431	$\frac{0.49}{0.65}=0.76=\frac{1.33}{1.76}$	$\frac{0.36}{0.48}=0.74=\frac{1.30}{1.76}$	. . .	$\frac{3.37}{3.65}=0.93=\frac{1.63}{1.76}$	$\frac{4.26}{6.55}=0.65=\frac{1.14}{1.76}$	. . .	*d*
. . . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
. . . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Heating and plumbing equipment	105,888	*d*	$\frac{0.17}{0.48}=0.35=\frac{0.23}{0.66}$	. . .	$\frac{4.54}{3.65}=1.25=\frac{0.82}{0.66}$	$\frac{5.01}{6.55}=0.76=\frac{0.50}{0.66}$	. . .	$\frac{9.38}{6.53}=1.44=\frac{0.94}{0.66}$
Structural metal products	284,121	$\frac{0.50}{0.65}=0.77=\frac{1.35}{1.76}$	$\frac{0.10}{0.48}=0.22=\frac{0.38}{1.76}$	. . .	$\frac{3.95}{3.65}=1.09=\frac{1.91}{1.76}$	$\frac{6.02}{6.55}=0.92=\frac{1.62}{1.76}$	. . .	$\frac{7.80}{6.53}=1.19=\frac{2.10}{1.76}$
. . . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
. . . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
Miscellaneous Manufactures	357,153	*d*	$\frac{0.57}{0.48}=1.20=\frac{2.64}{2.22}$	. . .	$\frac{5.64}{3.65}=1.55=\frac{3.43}{2.22}$	$\frac{6.67}{6.55}=1.02=\frac{2.26}{2.22}$	. . .	$\frac{12.62}{6.53}=1.93=\frac{4.28}{2.22}$

d indicates the unavailability of census data for computation.

is listed directly below the name of the state. Each row refers to an industry in the three-digit classification of the United States census; the total United States employment in each industry is listed in the corresponding cell of the first column. Each of the other cells of the table contains a pure number and two employment percentage ratios. The first ratio of percentages has as its *numerator* the given region's percentage share of total system employment in the given industry. (For example, the numerator of the first ratio in the second cell of the second row records the per cent of total United States employment in the meat products industry which is in Maine.) The first ratio of percentages has as its *denominator* the given region's percentage share of *all* manufacturing employment in the system. (For example, the denominator of the first ratio in the second cell of the second row records Maine's percentage share of total manufacturing employment in the United States.)[23]

When the first ratio of each cell in the body of the table is expressed as a pure number, we have the location quotient, as defined in section B, Chapter 5. Where the location quotient is less than unity, the given region has less than its "fair" share of the industry in question. Where the location quotient exceeds unity, the given region has more than a proportionate share of the industry in question.

But as has also been indicated in section B of Chapter 5, the location quotient is equal to a second ratio of employment percentages. The numerator of this second ratio indicates the per cent of the given region's total manufacturing employment accounted for by the given industry. (For example, employment in meat products manufacture accounts for 1.06 per cent of total manufacturing employment in Maine.) The denominator of this second ratio indicates the per cent of the over-all system's total manufacturing employment accounted for by the given industry. (For example, employment in meat products manufacture accounts for 1.93 per cent of total United States manufacturing employment.)[24]

2. THE COEFFICIENTS OF LOCALIZATION AND REDISTRIBUTION

With the systematic recording of data such as outlined in Table 7, the analyst is in a position to derive a number of useful descriptive coefficients and indexes. One that has been used extensively is the coefficient of localization.[25] This is a measure of relative regional concentration of

[23] It is to be noted that the denominator of the first ratio remains the same for any given column and only differs from column to column.

[24] Note that in any given row the denominator of the second ratio remains a constant; it varies only from row to row.

[25] See P. S. Florence [19], pp. 34ff; or P. S. Florence, W. G. Fritz, and R. C. Gilles [20], ch. 5.

a given industry compared to some total national magnitude such as population, land area, manufacturing employment, or income. It is essentially a comparison of the percentage distribution by region of employment in the given industry with the regional percentage distribution of the base magnitude, for example total national manufacturing employment. The actual computation of the coefficient typically consists of (1) subtracting for each region its percentage share of total system employment in the given industry (as recorded in the numerators of the first ratio in the cells of the given row of Table 7) from its percentage share of total manufacturing employment in the system (as recorded in the denominators

TABLE 8. DATA FOR COMPUTATION OF COEFFICIENT OF LOCALIZATION

Item	Regions			
	A	*B*	*C*	*D*
1. Per cent of employment of industry *i*	20	30	35	15
2. Per cent of total United States manufacturing employment	15	20	30	35
Difference (row 1 − row 2)	+5	+10	+5	−20
(Location Quotient)	(1.33	1.5	1.17	0.43)

of the first ratio in the cells of the same row of Table 7); (2) adding all positive differences, or all negative differences; and (3) dividing the sum of the positive (or negative) differences by 100. For example, if the data for a four-region system are as shown in Table 8, the coefficient of localization is

$$+20/100 = 0.2 \text{ (footnote 26)}$$

The limits to the value of the coefficient are 0 and 1. If the given industry is distributed exactly the same as is the base magnitude, the value will be 0. In contrast, if the entire industry is concentrated in one (small) region, the value will approach unity.

For the regional analyst seeking to implement a policy of diversification, a series of localization coefficients, each derived from the data of a relevant row in Table 7, could be useful. It could provide the basis for a preliminary and tentative judgment about which industries to seek and encourage or at least to investigate further. Industries with low coefficients are relatively

[26] The summation could just as well be of the minus deviations, since the percentage distributions are such that the sum of total plus and minus deviations is zero.

nonconcentrated regionally and are thus presumably amenable to location in a region seeking industrial diversification.

The basic feature of the localization coefficient—the comparison of two percentage distributions applicable to a given set of regions—can, of course, be extended to the comparison of any two meaningful percentage distributions. As already suggested, instead of using total manufacturing employment as the base, an analyst can use other magnitudes such as employment in another related industry or industrial complex, population, land area, Gross Product, and income. (The data may again be organized along the lines of Table 7.) If employment in another related industry is used as the base, the coefficient of localization is essentially the coefficient of geographic association, as defined by Florence.[27] It compares the geographic distribution of a given industry to the geographic distribution of the base industry. If population is used as base, the coefficient of localization may again be alternatively stated as a coefficient of geographic association whereby the geographic distribution of a given industry is associated with the geographic distribution of population.[28]

Not only are there many possible base magnitudes but also there are many magnitudes relevant for comparison with a base. That is, not only may tables like Table 7 be constructed in order to relate regional employment by industry to such base magnitudes as population, land area, and income, but they may also be constructed to relate to a pertinent base many other variables: for example, population by age group, color, or native stock; value added by industry; and urbanization by size class of city. Each such table then provides the data in the basic form for the

[27] When the value of the coefficient is zero, complete geographic association exists; when the value is unity, no geographic association exists.

[28] When in Table 7 a new base is substituted for total manufacturing employment, the row of numbers representing totals for the United States and its regions, which comes at the top of the columns, must be changed. And as a consequence the denominator of the first ratio, the pure number, and both the numerator and denominator of the second ratio must be changed.

For example, if population substitutes for total manufacturing employment as base, the population of the United States, and of each of its states, must be listed at the head of the respective columns. The numerator of the first ratio remains unchanged, since it represents a region's percentage share of employment in a given industry. The denominator of the first ratio changes; it now represents the region's percentage share of United States population. The numerator of the second ratio changes; it is the fraction formed by dividing a region's employment in a given industry by that region's population (which equals the region's per capita employment in the given industry). The denominator of the second ratio also changes; it is the fraction formed by dividing United States employment in a given industry by United States population (which equals United States per capita employment in the given industry). Also, as a consequence the pure number (location quotient) in each cell changes.

computation of the relevant set of coefficients of localization (or geographic association).[29]

One variant of the coefficient of localization which is of general value is the coefficient of redistribution. This coefficient is essentially a measure of the deviation between two distributions of the same phenomenon taken at different key points of time. For example, for two successive census years the percentage distribution of population by region could be compared. Taking one percentage distribution as the base, the deviations of the other percentage distribution can be computed. Summing all positive (or negative) deviations yields a figure which when divided by 100 can be designated a coefficient of redistribution. The value of such a coefficient will range from 0 (no redistribution) to unity (complete redistribution).

A number of other related coefficients may be constructed for various purposes. Some of these, together with the coefficient of geographic association and coefficient of redistribution, are listed in Table 9.[30] The relationships expressed by these coefficients and their possible uses are for the most part self-evident.[31]

[29] When in Table 7 a new nonbase magnitude is substituted for employment by industry, the first column of numbers representing totals for the several classes of industries must be changed. They must represent relevant totals for the new set of sectors (item classes or groups). And as a consequence the numerator of the first ratio, the pure number, and both the numerator and denominator of the second ratio must be changed.

For example, if the base magnitude is population (as in the previous footnote) and if number of families by income group substitutes for employment by industry as the nonbase magnitude, the number of families in each income group must be recorded in the first column. The numerator of the first ratio changes; it comes to represent a region's percentage share of the total number of families in a given-size income group in the United States. The denominator of the first ratio remains unchanged; it still represents the region's percentage share of total United States population. The numerator of the second ratio changes; it is the fraction formed by dividing the number of families of a given-size income group in a region by that region's population. The denominator of the second ratio also changes; it is the fraction formed by dividing the total number of families of the given-size income group in the United States by the population of the United States. Also, as a consequence, the pure number (location quotient) in each cell changes; it now reflects the extent to which the total population of each region has a proportionate share of United States families in a given-size income group.

[30] Also see P. M. Hauser, O. D. Duncan, and B. Duncan [30].

[31] As with the coefficient of localization, the coefficients of Table 9 are based on ratios of two percentages or fractions which in turn may yield the location quotient or one of its many possible variants. In connection with the use of the 1954 *Census of Manufactures* data, Alexander lists the following fractions (or percentages) of possible value for geographic analysis: number employed in manufacturing divided by total employed labor force, by total population, by total number of factories, or by employment in activity i; value added divided by total population or by number employed in manufacturing; value of payroll divided by value added; number of small factories divided by total number of factories (J. W. Alexander [4], pp. 20–26).

A series, by areas, of any one of these fractions when compared with the relevant fractions for the total system (say United States) could form the basis for a coefficient similar to those listed in Table 9. For example, the quotients of the per cent of a county's factories which are small factories to the per cent of United States factories which are small factories may be computed. The resulting series, by county, can be transformed into a series of ratios indicating a county's percentage share of all small factories in the United States to that county's share of all factories in the United States. By a summation of positive (or negative) deviations, a coefficient of geographic association of small factories relative to all factories can be computed.

3. THE LOCALIZATION CURVE

A tool superior in several ways to the coefficient of localization and related coefficients listed in Table 9 is the *localization curve*.[32] The localization curve is constructed from a set of regional percentage figures by plotting on the vertical axis a cumulative percentage figure for the given industry's employment and on the horizontal axis the corresponding

TABLE 9. COEFFICIENTS: TYPE *A*

Name of Coefficient	Author	Distributions Compared
Coefficient of geographic association	Florence et al. [20].	Shares of manufacturing employment by states: industry i versus industry j.
Coefficient of concentration of population	Hoover [36]	Shares by states: population versus area
Coefficient of redistribution	Hoover [36] Florence, et al. [20]	Shares of population (or total wage earners, or employment in selected manufacturing industries) by states: year α versus year β
Coefficient of deviation	Hoover [36]	Shares of population by states: white versus Negro
Index of dissimilarity	Duncan [14, 15]	Shares of workers by areas: occupation group *A* versus occupation group *B*
Index of segregation	Duncan [14, 15]	Shares of workers by areas: specific occupation group versus all other occupation groups

[32] This curve is developed and discussed in E. M. Hoover [35]. Also see E. M. Hoover [34], pp. 182–184.

cumulative percentage figure for the base magnitude. Typically, the required regional percentages can be obtained from the data included in a row of such tables as Table 7. [For every row (industry) of Table 7, a localization curve can be constructed.] The procedure involves (1) ranking regions by location quotients along the relevant row; and (2) plotting regions by rank on a cumulative percentage basis. For example, we may

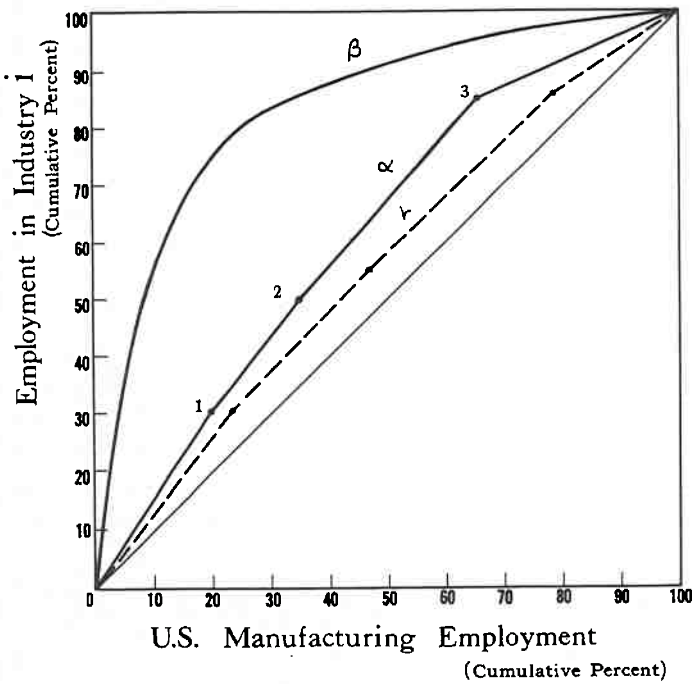


Figure 1. The localization curve.

take the data given in Table 8. Region *B* has the highest location quotient. As the first step, we therefore plot its percentages in Figure 1 (point 1 on curve α). Region *A* has the next highest location quotient. We therefore add region *B*'s percentages to the corresponding percentages of region *A* and plot the two sums (point 2 on curve α). Region *C* ranks third in size of location quotient for industry *i*. Its percentages are added to the corresponding sums already obtained and the resulting two new sums are plotted (point 3 on curve α). Finally, region *D*'s percentages are added to yield 100 per cent for both magnitudes. Joining the successive points by straight-line segments yields localization curve α.

Localization curves are essentially a device to depict and rank regions by location quotient since the slopes of their straight-line segments are identical with the location quotients of the several regions. If it turns out that a given industry is distributed regionally exactly the same as the base magnitude, the location quotients will all be unity and the localization curve will be a 45° diagonal from the origin. However, any divergence in the two distributions will be reflected in a deviation of the localization curve above and to the left of the diagonal. The extent of this deviation is a measure of the regional concentration of the industry, compared to the base magnitude. In this connection, we may compute the ratio of (1) the area between the localization curve and the diagonal to (2) the total area of the right triangle formed by the diagonal, the vertical axis, and the top of the graph. The limiting values of the ratio would be zero and one, as they were for the coefficient of localization computed from the plus or minus deviations of the percentage distributions.

In addition to using a localization curve to summarize the geographic pattern of an industry at a given point of time, an analyst may wish to proceed further with this tool. He may wish to contrast the geographic patterns of several industries at a given point of time. On a figure such as Figure 1 he may wish to construct another localization curve representing a second industry, still another localization curve representing a third industry, etc. For example, in Figure 1 a second localization curve β is constructed.[33] The advantage that a graphic presentation of two or more localization curves has over a presentation of two coefficients is clear-cut. This advantage exists whether or not the same set of regions is used in constructing the localization curve (and calculating the coefficient of localization) for each of the two or more industries under study.[34]

Moreover, an analyst may wish to compare the geographic pattern of an industry at a key point of time with its pattern for one or more other key points in time. On a figure such as Figure 1 he may wish to construct a localization curve for each time point to be considered. For example, on

[33] The β curve is taken from E. M. Hoover [34], p. 183, and represents shoe manufacture in the United States. States are taken as regions. The base magnitude is population. The nonbase magnitude is employment in shoe manufacturing.

[34] When different sets of regions are justifiably used from one industry to the next, considerable care must be exercised in reaching conclusions, whatever the tool used for descriptive comparison. See later remarks on the variation of the coefficient of localization with change in regional classification.

Figure 1 we have constructed localization curve γ which represents the same industry i as does curve α but is applicable to a different point of time. A comparison of curves α and γ, particularly when the ranking of regions is the same for both, has clear advantage over a mere presentation of two coefficients.

Thus, it may be concluded that the localization curve is a useful supplement (if not substitute) of the coefficient of localization. It retains regional detail in that the slopes of its line segments register the relevant regional location quotients, that is, show the regional components of a geographic

TABLE 10. CURVES AND COEFFICIENTS: TYPE *B*

Name of Coefficient	Author	Cumulative Distributions Compared	Order of Cumulation
Urbanization curve and coefficient	Hoover [36]	Shares by cities: employment in individual industry versus total population	By city size (small to large)
Urbanization curve and coefficient	Duncan [12]	Shares by city-size groups: retail sales in a given business versus total retail sales	By size of group (small to large)
Index of centralization	Duncan [15]	Shares by census tracts: specific occupation group versus all occupation groups (alternatively, employment in a given industry versus all industry)	By distance from city center

pattern. And it permits a visual comparison which for many studies may effectively complement (or replace) the presentation of one or more coefficients. Yet, for systematic studies which are based on a comprehensive set of tables, such as Table 7, and in which a fine industrial classification is employed, a complete set of localization curves may be an unwieldy tool for analysis. It may be much less efficient than the summary presentation of a set of coefficients, such as those listed in Table 9, particularly when supplemented by one or more sets of other coefficients, such as those listed in Table 10.[35]

[35] Hoover and others have attempted to extend the graphic technique of the localization curve to compare distributions ordered in accordance with other external criteria. For example, Hoover's urbanization curve is obtained by the same method as the localization curve, except that the units of the distributions are cities of various sizes rather than regions or states, and the order of the graphic cumulations is according to city size. Unlike the localization curve, the urbanization curve may be quite irregular or erratic. In fact, it may be above the diagonal in some places and below it in others. A numerical coefficient of urbanization can be computed as an area ratio, similar to the Hoover derivation of the coefficient of localization. Table 10 contains a summary list of a few such "externally ordered" curves and coefficients which are designated type *B*.

4. THE SHIFT RATIO AND RELATIVE GROWTH CHART

In section 2 we have already commented on the coefficient of redistribution. Another measure of regional shift in industry which is very similar to this coefficient is a *shift ratio*.[36] The rate of growth of employment in a given industry is first calculated over an intercensal period on an over-all or national basis. Then there is computed for each region the difference between the actual employment in the industry and the employment that would have resulted had the region's rate of growth in the industry been the same as the national rate. A positive difference signifies a shift of the industry into the region; a negative difference indicates a shift out of the region. The shift ratio for the industry is calculated by summing all the positive (or negative) shifts in employment and expressing the result as a proportion of total industry employment.[37]

It is apparent that an important defect of both shift ratios and coefficients of redistribution as measures of interregional industrial shifts is the fact that they take no account of changes in other major variables. Regional realignments in population, total income payments, value added by manufacture, private investment expenditure, public spending on waterways and highways, etc., may significantly influence or modify the possible implications of an industry's shift ratio or redistribution coefficient.[38]

[36] See D. Creamer [8], ch. 4. For a more recent application of this general type of thinking, see V. R. Fuchs [22].

[37] In addition to his use of shift ratios to measure regional redistribution of industrial employment, Creamer developed a rough measure of regional concentration of industry —the coefficient of scatter ([8], p. 90). It is expressed as the least number of states necessary to account for 75 per cent of total industry employment. Clearly, this can offer only a very general indication of the extent of industry concentration. More accurate comparisons are possible via the localization curve or coefficient of localization.

[38] For example, the fact that an industry has had a high shift ratio or coefficient of redistribution over a period of time may be considerably less striking if it is found that the coefficient of population redistribution was correspondingly high during the same period, particularly if there was a high coincidence in the individual components of the two types of ratios. Conversely, an industry with a low shift ratio or redistribution coefficient may be thought to have few implications of regional change—until it is found that the coefficient of population redistribution was high for the period. One possible way to attack the type of difficulty illustrated would be to compute a supplementary

One approach to this problem which can take account of one variable in addition to the industry change is the *relative growth chart*.[39] This is a graphic presentation of the scatter diagram type and can be adapted to the problem of industry redistribution using a figure such as Figure 2. In Figure 2, the vertical axis measures for a given industry employment at the

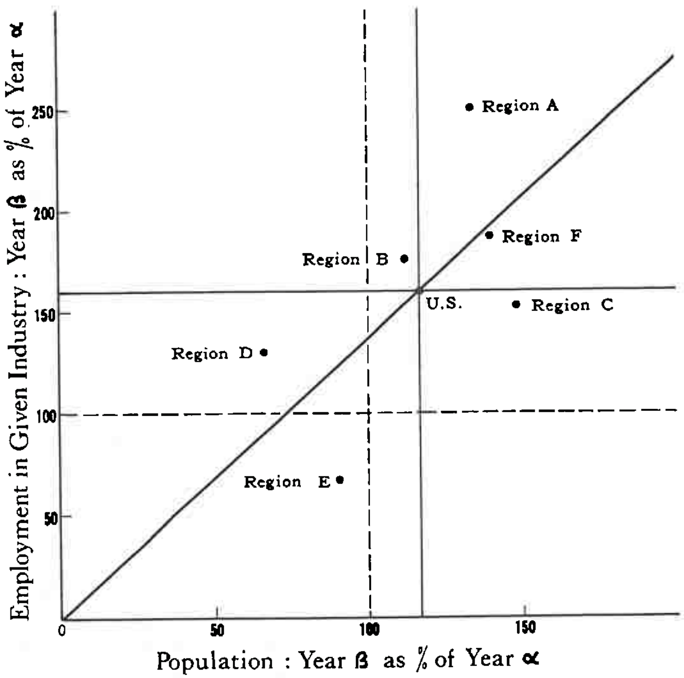


Figure 2. Relative growth chart, by region.

end of the period of analysis as a percentage of the employment at the beginning of the period. The horizontal axis measures a similar percentage for the other variable, say population. Each region of a system, here taken to be the United States, is represented by a point on the graph with coordinates determined by its percentage changes in industry employment and in population. The system's average percentage changes in

coefficient measuring the deviation between the regional percentage distributions of growth in population and growth in the given industry.

[39] See E. M. Hoover and J. L. Fisher [37], pp. 195–203.

these magnitudes are also indicated by a point, here indicating the average changes for the United States. A diagonal drawn from the origin through the United States point has a slope equal to the ratio of the two relevant percentages for the United States as measured on the two axes. This slope is also equivalent to the percentage change in United States per capita employment in the given industry.[40] The diagonal permits easy visual comparison of the several regional per capita changes with the United States per capita change. If a region is represented by a point lying above and to the left of the diagonal, its per capita change in the given industry employment is greater than that for the nation (system) as a whole. In addition to the diagonal, a vertical and a horizontal line may be extended from the axes through the point representing United States experience, as is done in Figure 2. These permit visual comparison of regional rates of both population change and employment change in the given industry with the corresponding United States rates.[41, 42]

[40] That is, if E represents employment, P population, and α and β the beginning and end of the period respectively, then

$$\frac{E_\beta/E_\alpha}{P_\beta/P_\alpha} = \frac{E_\beta/P_\beta}{E_\alpha/P_\alpha} = \frac{\text{per capita employment in year } \beta}{\text{per capita employment in year } \alpha}$$

[41] To illustrate the use of such a graph, a few of many possible situations may be hypothesized. An industry growing at somewhat the same rate as national population might show a wide scatter of points clustered along the diagonal. This would indicate that, although the industry had a high coefficient of redistribution, it had little change in *per capita* importance by regions. On the other hand, a wide scatter of points along the horizontal line would indicate considerable divergence among the regions in *per capita* changes in the given industry, in spite of a low coefficient of redistribution. If, however, there had been a situation of major regional redistribution of the industry combined with little relative regional change in population, the result would be a wide scatter of points all close to the vertical line. Not only would the industry show a high redistribution coefficient; it would also show a wide extent of change in its regional per capita importance.

[42] For regional analysis, the relative growth chart is useful in ways other than indicated in the text. To cite two such ways, it can compare by regions (1) per cent changes in income, in per capita income, and in population; or (2) per cent changes in Gross Product, in productivity per worker, and in total employment.

Also, in line with certain suggestions of Zelinsky, the influence of population change could be accounted for directly in computing the individual ratios of the redistribution coefficient (W. Zelinsky [69]). In measuring change in United States manufacturing activity between 1939 and 1947, Zelinsky develops certain "factors" or expressions which take account of the interrelations between areal changes in amount of manufacturing, population, and number of production workers. For example, he writes the $V:P$ factor for an area as

$$V_{1947} - \frac{V_{1939} \cdot P_{1947} \cdot K}{P_{1939}}$$

where V is value added by manufacture, P is area population, and K is a constant such

5. TECHNICAL LIMITATIONS

Thus far, we have suggested some possible uses for the various coefficients and related concepts which have been discussed. However, like most techniques, they are subject to major limitations. One of the most evident shortcomings of any coefficient or graphic representation which is based on the deviation between, or ratio of, two percentage distributions is that the results obtained will differ, depending on the degree of areal subdivision. For example, the coefficient of localization of an industry compared to total manufacturing workers would almost certainly be higher if the nation were broken down by counties rather than by states.[43] Furthermore, the degree of variation in the value of the coefficient under such conditions would differ for different industries. Thus, two industries might have virtually the same coefficient of localization if states were the unit of subdivision but substantially different ones if counties were. This reduces the usefulness of interindustry comparison based on the coefficient of localization and similar devices.

This shortcoming is neatly portrayed by Figure 3, taken from Duncan, Cuzzort, and Duncan.[44] This figure indicates indexes of population concentration, for five alternative systems of areal subdivision of the United States, 1900–1950. These indexes are essentially coefficients of localization of population, where land area is the base magnitude. That these coefficients decrease as the size of region increases is clear from this figure. For any one year, the smaller the areal subdivision the greater the coefficient proves to be. More striking, however, is the fact that over the time period examined the coefficients based on large areal subdivisions

that the sum of the $V\!:\!P$ factors for all areas is zero (i.e., $\sum V\!:\!P = 0$). (K thus approximates the ratio for the nation of per capita value added by manufacture for 1947 to the same magnitude for 1939.) Thus, if the ratio of per capita value added had increased at a uniform rate in all areas between 1939 and 1947, the $V\!:\!P$ factor for each area would be zero.

Although Zelinsky presents his findings as a set of positive and absolute differences (see Map 2), they can be expressed in terms of percentile differences. Such percentile differences in turn could form the basis for the calculation of a redistribution coefficient as already discussed. Use of magnitudes other than value added and population would also lead to other relevant redistribution coefficients.

43 Thompson points out that virtually any industry exhibits a high coefficient of localization if the areal subdivision scheme is fine enough. However, in his view, of more significance may be the rate at which the value of the coefficient decreases as larger subdivisions are considered. A rapid rate of decrease suggests that the industry is in reality rather dispersed, with the several sites (or areas) of production contiguous with areas of nonproduction. A slower rate of decrease indicates that the production sites are "clustered" within a smaller number of separate producing areas. See W. R. Thompson [64].

44 O. D. Duncan, R. P. Cuzzort, B. Duncan [13].

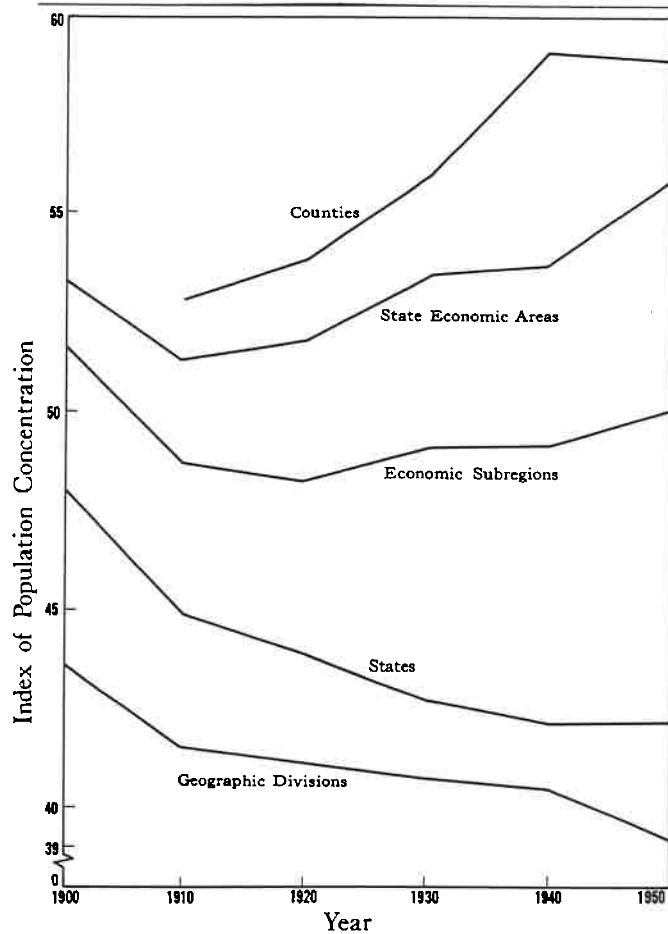


Figure 3. Indexes of population concentration, for various systems of areal subdivision of the United States: 1900 to 1950. Source: O. D. Duncan, R. P. Cuzzort, and B. Duncan [13].

tend to become smaller, whereas the same coefficients based on small areal subdivisions tend to become larger. This fact corroborates the point that as a descriptive device any given coefficient can be meaningful *only* with reference to the set of areal subdivisions adopted. Furthermore, this fact suggests that a series of coefficients based on different areal subdivisions is necessary to indicate the complex pattern of population changes over time.[45]

A second major difficulty of the coefficient of localization and related concepts reflects the tendency of any such measure to vary considerably, depending on the choice of base. Localization, centralization, redistribution, etc., are necessarily expressed relative to a base magnitude—there is no absolute measure. Thus, if a large portion of a country's total industry is concentrated in a relatively few metropolitan areas, a specific industry also heavily concentrated in these same areas will quite likely show a low coefficient of localization when the coefficient is computed with total industry employment or output as a base. If the coefficient were computed with geographic area as a base, the value would be considerably higher.[46]

[45] For further discussion, see O. D. Duncan, R. P. Cuzzort, and B. Duncan [13].

In view of this limitation, an analyst may seek measures of geographic distribution which are essentially independent of the scheme of subdivision used. One such measure is a centrographic technique which has been developed by Bachi for summarizing the extent of population dispersion. (See R. Bachi [5], as reported by O. D. Duncan, R. P. Cuzzort, and B. Duncan [13]). In a population distribution each areal unit, with a population of P_i, can be approximated as a point with a horizontal coordinate x_i and a vertical coordinate y_i. Then the "mean center" of population is at $(\bar{x}, \bar{y})$ where

$$\bar{x} = \sum_i P_i \cdot x_i \Big/ \sum_i x_i \quad \text{and} \quad \bar{y} = \sum_i P_i \cdot y_i \Big/ \sum_i y_i.$$

Dispersion around the mean center can be measured by the "standard distance,"

$$d = \sqrt{\frac{\sum_i P_i \cdot (x_i - \bar{x})^2 + \sum_i P_i \cdot (y_i - \bar{y})^2}{\sum_i P_i}}$$

Similarly, the dispersion of population can be measured in terms of distances separating the centers of the individual areas. There is a constant relationship between these two measures of dispersion, and their values are affected only incidentally by the type of subdivision employed. (Generally speaking, the measures are more precise the smaller the subdivisions.)

Clearly, such measures could be of considerable help in evaluating the degree of concentration or dispersion of given industries, number of manufacturing workers, etc. However, as Duncan, Cuzzort, and Duncan point out, when one is concerned with changes in a distribution pattern (e.g. over time), it is hardly likely that any single centrographic technique could furnish a complete or adequate description. A series of coefficients based on different sets of areal subdivisions might well be preferable.

[46] W. R. Thompson [64] has pointed out this problem of the implicit weighting of the individual regions by their respective shares of the base magnitude. In order to weight

As already noted, a defect which applies particularly to the numerical coefficients and shift ratios is that they express a combined or net value and give no indication of the behavior of the individual components making up that value.[47] In this connection, preserving the detail on the behavior pattern of the individual regions via the use of a localization curve would be highly desirable, although perhaps cumbersome.

As already pointed out in a previous footnote, the graphic approach has been extended to the construction of curves and the calculation of ratios and coefficients in which the distributions are ordered, not according to the magnitude of the individual deviations but according to some external standard, such as size of city, distance from city center, etc. Coefficients and curves of this type (type *B*) can be of only limited use. The curves are not typically smooth or symmetric and can give only a very rough idea of what they purport to describe. They can be badly distorted by major deviations anywhere in the ranking. The corresponding coefficients based on ratios of graphic areas are noncomparable.[48]

A final major difficulty of the coefficient of localization and related concepts is the problem of designing a proper set of industrial categories, income classes, occupational groups, population sectors, etc. Thus far we have assumed that a set of categories is predetermined, for example, that the most desirable industrial classification has been determined. Or

regions equally (when such is desirable), he suggests that a coefficient of spatial variation be substituted for the coefficient of localization. If we let E_j^L be employment in the given industry j in region L (where regions are numbered from 1 to U), E^L be total manufacturing employment (or population) in region L; h_j^L equal E_j^L/E^L, and U be the number of regions, then the coefficient of spatial variation equals $\sigma/\bar{h}_j$ where

$$\sigma = \sqrt{\frac{\sum_L \left(h_j^L - \bar{h}_j\right)^2}{U}}$$

and

$$\bar{h}_j = \frac{\sum_L h_j^L}{U}$$

[47] For example, a specific value for the coefficient of redistribution of employment in a given industry may have resulted from a major exodus from one region and minor increases in most of the others; or from a major expansion in one region coupled with small losses in the others. The knowledge of which situation (if either) actually existed would certainly be helpful to a regional analyst concerned with future prospects of the industry, but the coefficient itself would not furnish this information.

[48] For example, suppose that the urbanization curves of industry i and industry j enclose equal areas between the curve and the diagonal, giving the two industries equal coefficients of urbanization. However, if the curve for i has a bulge near one end of the ranking, and the curve for j has a similar bulge toward the other end, we could not state that industries i and j are equally "urbanized."

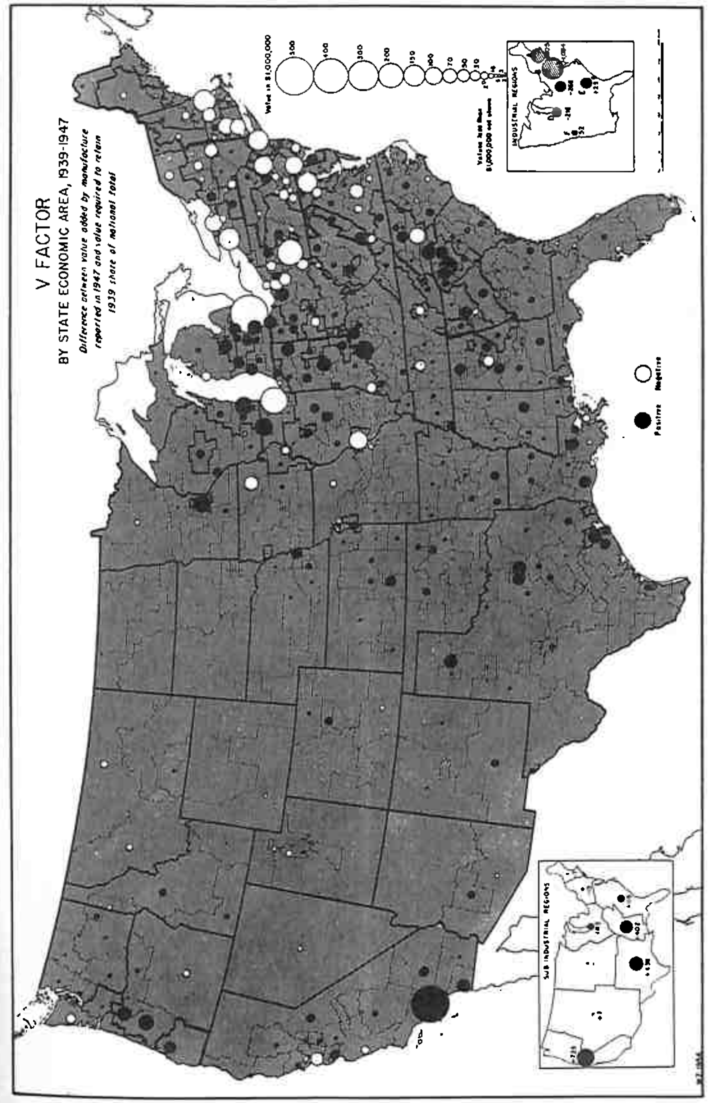
we have been concerned with a given industry (already defined), or manufacturing industry as a whole, etc., and have probed into its geographic pattern. But suppose our concerns center around broad issues such as resource development policy and industrial diversification within a system of regions. Suppose we wish to select, out of the whole array of industries, a few that *initially* appear suitable for development for each of a number of areas. Suppose, too, for this task we judge that there ought to be constructed a complete set of coefficients of localization, ratios, etc. based on systematic sets of data such as those contained in Table 7. Unfortunately, the values of the coefficients, ratios, etc., obtained will be very much dependent on the fineness of the industrial classification employed. A gross industrial classification, such as a two-digit one, for example, would tend to yield low coefficients of localization, etc., just as large geographic divisions do. In contrast, a fine industrial classification, such as a four- or five-digit one, would tend to yield high coefficients just as small areal subdivisions do. And, as will be evident in the next section, the ranking of regions by degree of specialization may be greatly influenced by the nature of industrial classification. Further, the pattern of change in these coefficients over time may be very much a function of the degree of industrial disaggregation.

6. CONCEPTUAL LIMITATIONS AND SUMMARY REMARKS

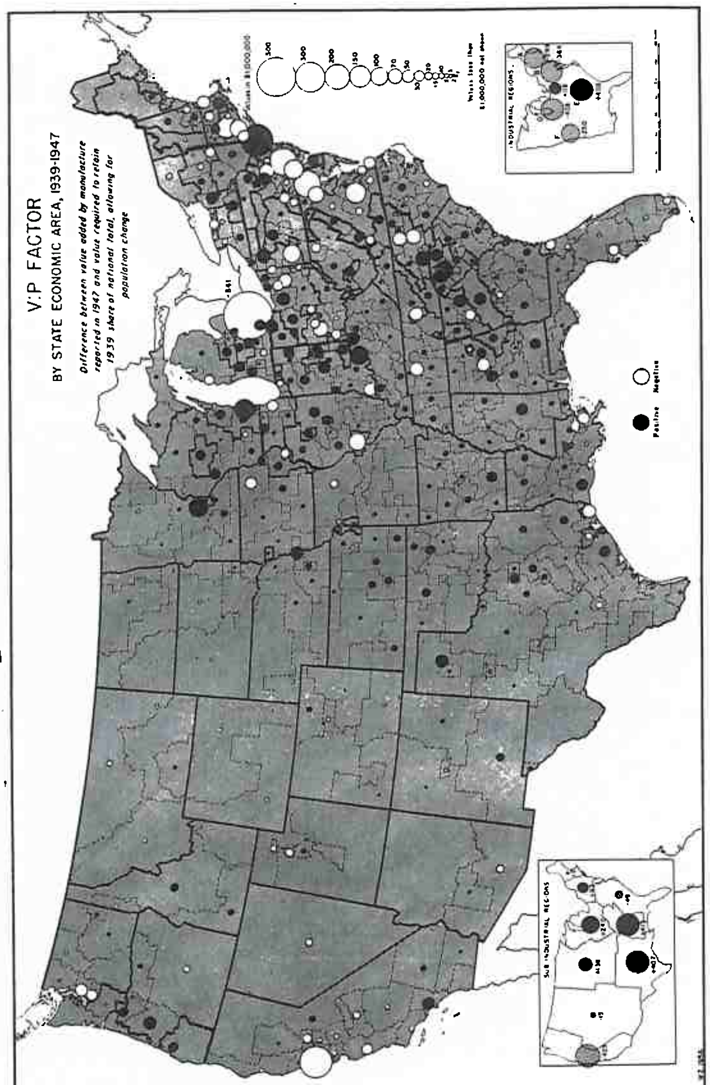
The limitations and defects discussed to this point are technical. They are direct consequences of the method by which the coefficients, ratios, and curves are defined or derived and data and regions classified. A more serious and fundamental limitation to their use is that they are of little help in identifying cause and effect relationships. They are essentially mechanical devices with which empirical facts can be processed to reveal certain statistical tendencies or regularities.

For example, consider Map 1. This map effectively presents basic data which might be used to develop a shift ratio, or a coefficient of redistribution. It portrays by State Economic Areas (S.E.A.'s) differences between value added by manufacture in 1947 and the amounts required to retain 1939 shares of national total value added by manufacture. (This is defined by Zelinsky as the *V* factor.) It shows, for example, that over the period 1939–1947, the shares of S.E.A.'s in the Middle Atlantic states and New England generally declined, whereas those of the S.E.A.'s in the Ohio–Indiana–Michigan region generally increased.

Since during this same period population growth of the S.E.A.'s has also varied considerably, another type of map such as Map 2 may be considered more relevant. Map 2 presents the value added changes of Map 1 after they have been adjusted by the population changes. (This set of



Map 1. *V* factor, by state economic area, 1939–1947. Source: W. Zelinsky [69], p. 110.



Map 2. *V* : *P* factor, by state economic area, 1939–1947. Source: W. Zelinsky [69], p. 116.

adjusted changes, which is defined by Zelinsky as the *V*:*P* factor,[49] is closely related to the localization curve, location quotients, and the coefficient of localization.) Map 2 indicates the same general pattern of change as does Map 1, but certain significant modifications can be noted. There are significant exceptions to the generally declining shares experienced by S.E.A.'s along the northern Atlantic Seaboard. Certain strategic areas in western New York State and in western Pennsylvania are associated with only *slightly* decreased shares.[50]

Clearly, maps such as these are extremely effective in establishing trends and patterns of change. Yet, it must be remembered that neither maps nor the corresponding coefficients explain or identify the economic and other forces which interact to produce these tendencies and regularities.[51] As a consequence, the current general trends and patterns revealed by the various curves and coefficients cannot be assumed to apply automatically to future development or, by analogy, to individual regional situations. This is not to deny that the various coefficients are valuable to the regional analyst as an aid in ordering and classifying his empirical data and in deciding which avenues of further research are likely to be fruitful. However, the definite limitations of the measures should be understood, and they should not be considered as "short cuts" to conclusions that can only result from more basic analysis.

The discussion of the general type of statistical measure exemplified by the localization curve, the shift ratio, and the coefficient of localization can be briefly summarized:

1. The coefficient, curve, or ratio is derived essentially from a comparison of two percentage distributions which have common units of classification, for example, states, counties, cities, census tracts, etc.[52] This formulation results in three important technical limitations. First, a change in the degree of fineness of area classification will generally cause a change in the coefficient, curve, or ratio. Second, the value of the coefficient, or ratio, or the shape of the curve is relative; it describes a

[49] This factor is defined in footnote 42.

[50] A relative growth chart could be constructed as an alternative method of presenting the data of Map 2. On the vertical axis would be measured value added in 1947 as a percentage of value added in 1939. On the horizontal axis would be measured population in 1947 as a percentage of population in 1939. Thus each region, and the United States as a whole, could be represented by a point on the relative growth chart with coordinates determined by the relevant percentages. Comparative analysis could then proceed as sketched in the text.

[51] For a detailed explanation of this basic limitation as it applies to one particular measure, the coefficient of geographic association, see Robert E. Kuenne [45], ch. 2.

[52] The relative-growth (Hoover-Fisher) chart is more flexible and permits comparison of per cent changes of three magnitudes, although one of the three is not independent.

given distribution in terms of a base distribution and is only as good as the base is relevant. Third, the value of the coefficient, or ratio, or the shape of the curve will tend to vary, depending on how broadly the nonbase magnitude (e.g., industry sector, income class, and occupation group) is defined.

2. As is true of virtually all statistical measures, the devices and concepts discussed are of little value in identifying or evaluating cause and effect relationships. They can assist the analyst to perceive certain general empirical associations but can be considered only as rough guideposts for basic regional analysis and planning.

E. Coefficient of Specialization, Index of Diversification and Related Concepts

Closely associated with the concepts discussed in the previous section are the coefficient of specialization, index of diversification, and related concepts.[53] Objectives similar to those mentioned at the start of the previous section have motivated the development of these latter tools and concepts. Also, these tools and concepts are based on data similar to those discussed in the preceding section.

To point up these interconnections, we re-examine Table 7. There we noted that each cell was made up of two ratios, each equivalent to the location quotient (the pure number) recorded. These two ratios are obtainable from one another simply by carrying through the algebraic operation of substituting for one another the denominator and numerator of the nonbase and base precentages, respectively.[54] We have already discussed how the set of the first ratios in a given *row* and the location quotients along a given *row* can be used to develop coefficients of localization and redistribution, localization curves, etc. If we now concentrate on the ratios (in particular the second ratio of each cell) and the location quotients *by columns* (i.e., by regions), we can derive the several tools and concepts to be discussed in this section.

As already noted, the numerator of the second ratio of a cell in Table 7 indicates for the given region (at the head of the column) the per cent of the employment of a region accounted for by the industry in the row of the cell, whereas the denominator of the same ratio indicates for the entire system (the United States) the per cent of its total manufacturing employment accounted for by the same industry. Paralleling the discussion of

[53] Because of such association, we shall not treat in this section certain fine points which have already been treated in the previous section. The reader interested in such points should read this section in parallel with the previous section.

[54] See footnote 3, Chapter 5.

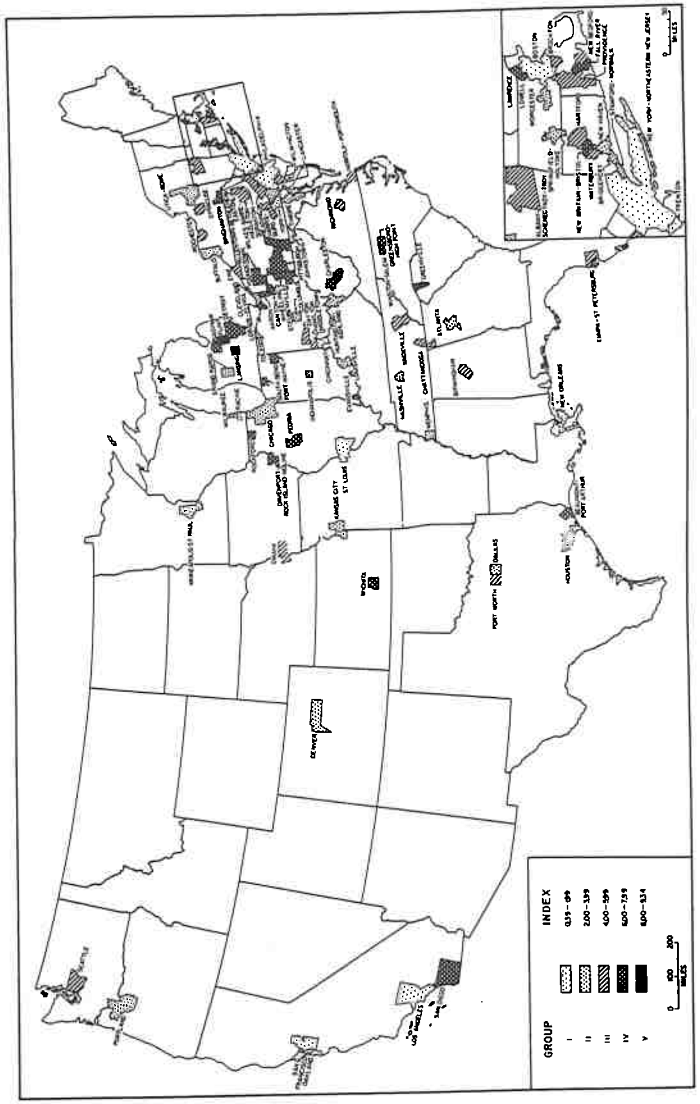
previous sections, we may compute a coefficient comparable to the coefficient of localization. We call this new coefficient, which pertains to a given region, *the coefficient of specialization* of that region. This coefficient is computed for the given region by: (1) subtracting the numerator from the denominator of each of the second ratios in the region's column; (2) adding all positive (or negative) differences; and (3) dividing the sum (without regard for sign) by 100. The limits to the value of this coefficient are 0 and 1. If the region has a proportional mix of industry identical with the system (United States), the coefficient will be 0. In contrast, if all the employment of the region is concentrated in a single industry, the coefficient will approach unity. This coefficient thus measures the extent to which the distribution of employment by industry classes in the given region deviates from such distribution for the United States. As with the coefficient of localization, this coefficient is helpful to the regional analyst seeking to implement a policy of diversification.

The basic feature of the specialization coefficient—the comparison of two percentage distributions applicable to a given set of classification units (e.g. industries, in Table 7)—can be extended to the comparison of any two meaningful percentage distributions for a given region versus the United States. For example, the percentage shares of total regional income accounted for by the members of each of the several income groups in a region can be contrasted with the corresponding percentage shares of national income accounted for by these same income groups. Or the percentage share of total regional employment accounted for by members of each occupational group in a region can be contrasted with the corresponding percentage shares for the nation. And so forth.[55]

When coefficients of specialization have been obtained for a number of regions, it is often helpful to map the coefficient values in order to point up contrasts among the regions. Such a map would resemble Map 3, which has been developed by Rodgers and which refers to values along an index of industrial diversification, a concept similar to the coefficient of specialization.[56]

[55] Comparisons such as the ones noted may involve changes (with respect to the data of Table 7) not only in the base magnitude and the nonbase magnitude, but also in the classification units to which the distributions of base and nonbase magnitudes apply. Consequently, any given comparison may involve changes in any of or all the percentages and ratios which appear in Table 7. See the discussion in footnotes 28 and 29.

[56] A. Rodgers [56, 57]. Rodgers' map depicts what he terms the refined index of diversification for each of a large number of metropolitan industrial areas. The refined index of diversification for an area is derived from the area's crude index of diversification, which crude index is computed as follows. Percentages of total area employment in each of 22 manufacturing groups are calculated. These percentages are ranked in order from highest to lowest. Then the percentages are cumulated, one at a time, to



Map 3. Index of industrial diversification by United States metropolitan regions. Source: A. Rodgers [56], following p. B-7.

Corresponding to the coefficient of specialization for a region is a *specialization (or diversification) curve*. Such a curve is constructed in essentially the same manner as the localization curve. The vertical coordinates of successive points on the curve measure cumulative percentages, industry by industry, of the region's total manufacturing employment. The corresponding horizontal coordinates measure cumulative percentages, industry by industry, of total United States manufacturing employment. The industries are ordered according to the value of the given region's location quotients for the industries, as recorded in a column of such tables as Table 7. The ordering is from largest to smallest. For any given region, the deviation of the curve from a diagonal from the origin will measure the degree to which the distribution among industries of the region's manufacturing employment differs from the corresponding distribution of United States manufacturing employment. A variant of the coefficient of specialization could be derived from the specialization curve by computing the ratio of (1) the area between the specialization curve and the diagonal, to (2) the total area of the right triangle formed by the diagonal, the vertical axis, and the top of the graph. The limiting values of this ratio will be zero and one, as they are for the coefficient of specialization computed from the plus or minus deviations of the percentage distributions.[57]

yield a set of cumulative subtotals, that is, the largest is set down first, then the sum of the largest and the next largest, then the sum of the largest and the next two largest, etc. Summing these cumulative subtotals yields the area's crude diversification index. If all the employment of an area were concentrated in one manufacturing group, the area's crude index of diversification would be 2200. This figure would represent the crude index value for least diversity. In contrast, if employment were equally distributed among the 22 manufacturing groups, the area's index value would be approximately 1150, a value representing greatest diversity.

The refined index of diversification for an area, as defined by Rodgers, is equal to (1) the area's crude index minus the crude index for all industrial areas taken together, divided by (2) the crude index for least diversity minus the crude index for all industrial areas taken together. Thus, a refined index of zero for an area would indicate the same degree of diversification for that area as for all areas taken together. A value of +1.0, on the other hand, would indicate complete nondiversification.

The similarity of the refined index of diversification to the coefficient of specialization is apparent. An advantage of the index, however, is that it would take a negative value for areas which had a more even or equal distribution of employment among manufacturing industries than the over-all system of areas. The coefficient, on the other hand, measures only the degree of *deviation* of an area from the diversification pattern of the over-all system, whether that deviation is in the direction of more or less even distribution.

[57] A curve could be constructed based on an industry (or other) distribution ordered in accordance with some criterion other than size of location quotient. For example, the order could be based on the number of employees in each industry. The resulting curve would be irregular in shape, and ordinarily its usefulness would be greatly circumscribed.

In addition to utilizing a specialization curve to summarize the industrial diversification of a given region, the analyst may wish to achieve an interregional comparison by means of a set of these curves. For a given point of time he may plot on the same graph a specialization curve for each of

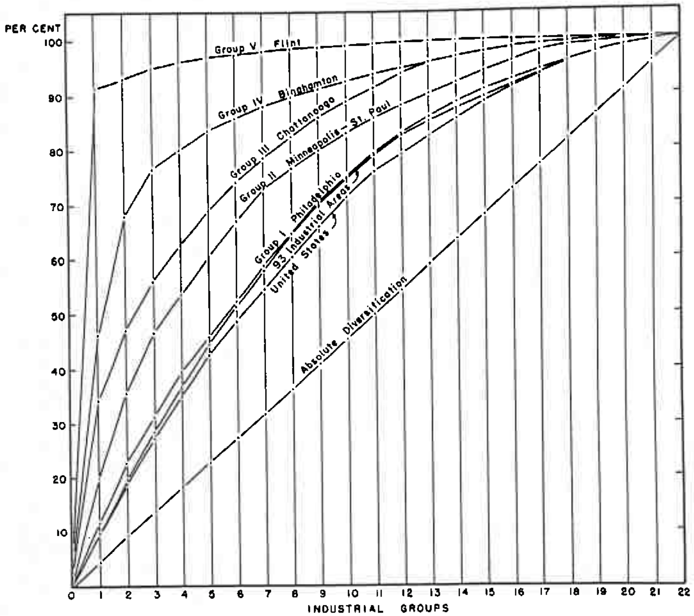


Figure 4. Crude diversification patterns, 1950; by selected metropolitan regions and the United States. Source: A. Rodgers, [56], following p. B-6.

several regions. Such a group of curves may resemble those of Figure 4.[58] Considerable insight into the comparative industrial structures of the regions of a system can be gained by a careful study of such a figure.

58 It is evident from Figure 4 that Rodgers' curves are constructed in a fashion slightly different from the way specialization curves (as defined) would be derived. Instead of comparing a region's distribution of manufacturing employment by industry with the corresponding total system distribution, Rodgers compares the individual region's distribution with an hypothetical equal distribution of employment among all industries. The latter is considered by Rodgers as "absolute diversification." The use of this latter curve has the advantage of providing an absolute point of reference with which to compare not only individual regions but also the system taken as a whole.

Moreover, a group of specialization curves can be plotted, each representing the pattern of specialization for a given region at a different point of time. Such curves can helpfully guide and facilitate the analysis of historical changes in regional diversification patterns.

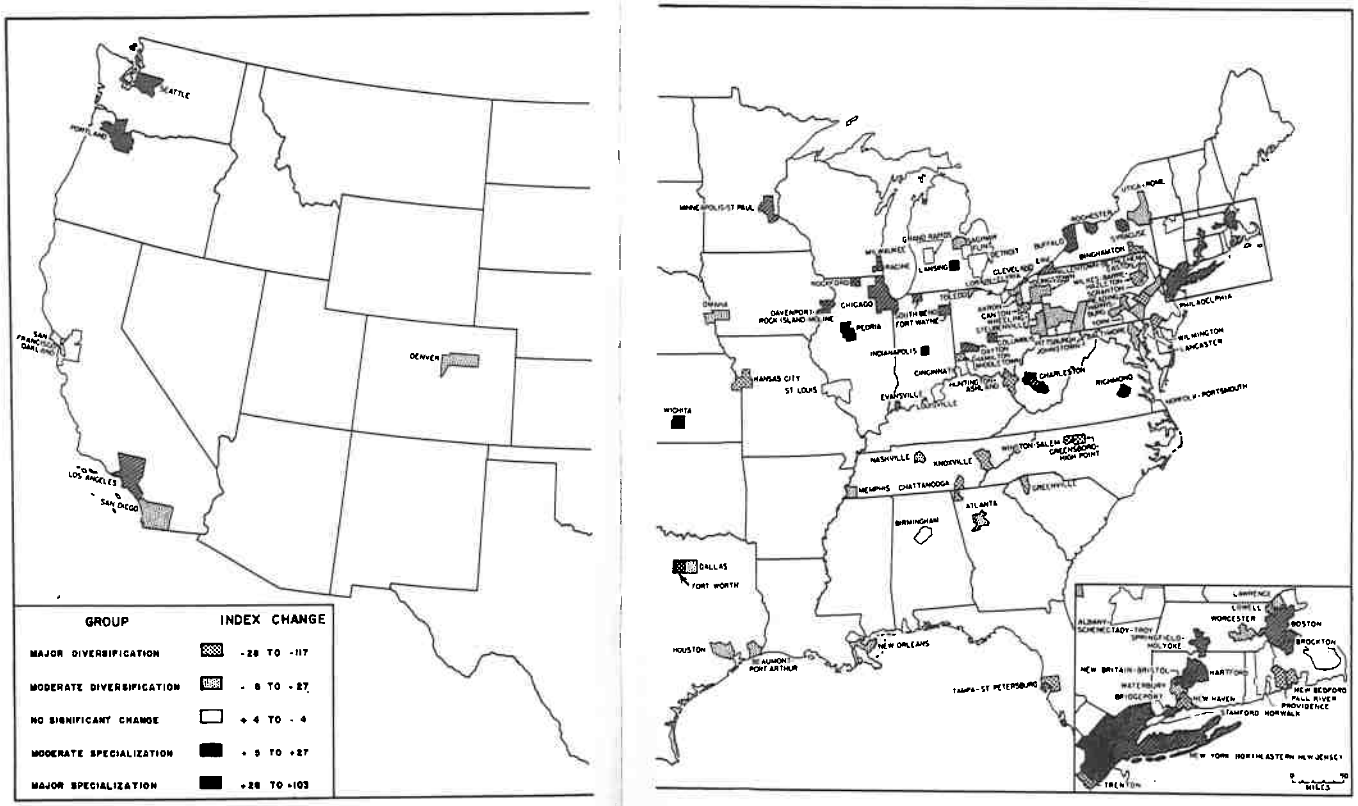
In many instances the specialization curve (or group of such curves) has clear-cut advantages over the corresponding coefficient(s) of specialization. Perhaps of principal importance is the fact that the curves give some indication of the relative contribution of the individual industries or industry groups to over-all diversification. This distinction may be particularly useful when several regions or several time periods are being compared.

As already indicated, we can compute for a given region a series of specialization coefficients over time as well as plot a series of specialization curves over time. In addition we may wish to summarize such analysis for the whole system of regions in order to observe the over-all change in specialization over time within the system. This problem can be attacked by (1) computing for each region the difference between its coefficients of specialization at two successive points of time; (2) summing over all regions; and (3) dividing by the number of regions. However, in view of the technical shortcomings of the coefficient of specialization, such a summarization will possess only a limited degree of validity and usefulness. Of much more use in this connection is a map that records changes in coefficients. Although such a map is not available, a similar map, constructed by Rodgers, on changes in the crude diversification index, 1940–1950, illustrates this point well.[59] This map is reproduced here as Map 4.

Corresponding to the coefficient of redistribution discussed in the preceding section, which coefficient summarizes the change over time of the regional distribution of some magnitude (e.g., population, industry employment, total manufacturing employment, etc.), *a coefficient of redistribution* within a region over time can be computed. This latter coefficient may relate to employment by industry, employment by occupation group, income shares by income group, etc. For example, the percentage distribution in a region of employment by industry group can be compared for any two successive census years. The resulting coefficient of redistribution based on differences of corresponding percentages will indicate the extent to which on a relative basis interindustry shifts of employment have taken place in the region during the intercensal period.

Similar results can be obtained by calculating an interindustry *shift ratio* for the region. The over-all rate of growth of the region's industrial employment can be calculated for the intercensal period. Then for each

59 A. Rodgers [56].



Map 4. Changes in crude diversification index 1940–1950 by metropolitan regions. Source: A. Rodgers [56], following p. B-10.

industry in the region there can be computed the difference between the actual employment in the latter census year and the employment that would have resulted had the industry's employment grown at the same rate as the region's total industrial employment. A positive difference will indicate a relative shift of employment into the industry, a negative difference a relative shift out of the industry. The actual shift ratio will be calculated by summing all the positive (or negative) interindustry shifts in employment for the region and expressing the result as a proportion of the region's total industrial employment.

It is evident that for some purposes a measure which can compare interindustry shifts of employment over time in a given region to the corresponding shifts for all regions will be more useful than just the coefficient of redistribution or shift ratio for the given region. One such device is a type of *relative growth chart*, illustrated by Figure 5. The vertical axis of this figure measures, for region A, employment in year β as a per cent of employment in year α, either for a single industry or for industry as a whole. The horizontal axis measures for the system (United States) the same percentage. Along the vertical axis point M represents this percentage for industry as a whole in region A. Along the horizontal axis point R represents the same percentage for industry as a whole for the

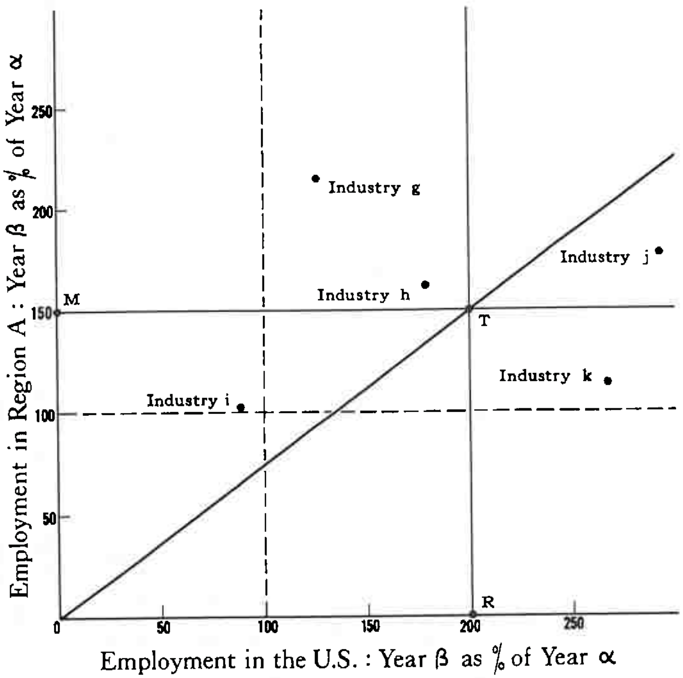


Figure 5. Relative growth chart, by industry.

United States. Therefore, the slope of the diagonal from the origin through point T (whose coordinates are OM and OR) measures the ratio of these two percentages. This slope is also equivalent to the ratio of region A's percentage share of total system (United States) employment in year β to region A's percentage share of total system employment in year α. The steeper the diagonal, the faster has been the growth in the region's total industrial employment compared to the United States rate of growth.

The diagonal of Figure 5 permits interesting comparisons for any particular industry of its growth in a given region relative to its growth in the total system. For example, the position of industry i in Figure 5 indicates that this industry is slow-growing. Additionally, although this industry's growth in region A was slower than average industrial growth in both this region and the United States, its relative decline in region A was less severe than in the system as a whole. Thus the industry fared better in region A than could have been expected on the basis of its performance in the United States. Or, to take another example, we may examine industry j. Its position on Figure 5 indicates that it is a fast growing industry. Its growth over the period exceeded average industrial growth in both region A and the United States. However, its growth performance in region A was less than could have been expected on the basis of performance in the total system. Hence, region A's share of this industry declined relative to region A's share of total industry.

This discussion of the relative growth chart completes the summary presentation in this section of concepts relating to specialization and diversification which parallel the concepts relating to localization covered in the preceding section. Because of this parallel, virtually the same general technical and conceptual limitations apply to the concepts presented in this section as apply to those in the previous section. Hence, it would be only repetitious to undertake a critical evaluation at this point. Suffice it to say that the values of the specialization coefficients and the shapes of the related curves are dependent on the degree of fineness of the units of classification as well as on the size of regions. More important, the coefficients, curves, and other devices discussed in this section are essentially descriptive and cannot identify cause-effect relationships. They, too, take on meaning only when embraced by a valid conceptual and theoretical framework.

F. General Conclusions

In earlier chapters we have not explicitly injected into regional analysis the major factor of optimizing behavior. This factor underlies the comparative cost approach. It lends to this approach a causative significance, of considerable validity for predictive purposes. In the orthodox single-industry framework, the comparative cost approach has had a widespread and fruitful application. And because of its optimizing rationale, in the future it promises to be one of the most powerful tools in the kit of the regional scientist. Yet it is recognized that this approach in its traditional framework pertains to partial equilibrium only; that is, comparative cost analysis looks at behavior within a single industry, the structure of all other industries as well as demand, prices, and costs being assumed as given. Therefore, as already noted, sound regional analysis requires that the traditional comparative cost approach be supplemented with more general techniques capable of cutting through the restrictive bounds of single-industry analysis. This supplementation will be pursued in the subsequent chapters dealing with interregional and regional input-output techniques, industrial complex analysis, interregional linear programming, and gravity models. Additionally, factors outside the economic sphere must be weighed in forming locational decisions, and some developing techniques which attempt to appraise such factors are sketched in Appendices A and B.

Surrounding the comparative cost approach are a number of measures relating to industrial location and regional distribution of phenomena. Some of these measures—the labor coefficient and others discussed in section C—are very useful in preliminary stages of research. They represent incomplete comparative cost ratios, which in a general way short-cut analysis. Full analysis, however, requires that they be converted into comparative cost ratios. Other measures—coefficients of localization, localization curves, shift ratios, indexes of specialization, and the several other measures discussed in sections D and E—are essentially designed to describe and summarize systems of industrial locations, population and subpopulation locations, and locations of other relevant items and phenomena. They are valuable for portraying the "what" of systems as they are or have been. They permit a view of the internal structure of regions along several dimensions. They permit the comparison of a given region's structure with the structure of other regions and, where justifiable, with the system as a whole or other fictitious norms. Moreover, they permit identification of changes over time in the structure of both the region and the system. Thus, in a very important way they supplement approaches such as comparative cost analysis which requires, both for implementation and testing, structural knowledge in the form of factual materials on the outcome of the interplay of underlying forces, both currently and in the past.

It is to be noted, however, that such supplementation is not of an analytical nature. For example, although comparative cost analysis can suggest changes to be expected in coefficients of localization, shift ratios, specialization curves, etc., the converse cannot be stated. True, a high coefficient of localization for a particular industry may reflect a major cost differential or scale economy factor, but of itself it tells nothing of what the factor is, how important it is in relation to others, and to what extent this factor is expected to persist. Likewise, when the coefficient of localization and other related measures are used in conjunction with population and migration estimation, income and balance of payments statements, and commodity and money flow studies, they add little to the understanding of basic interrelations and to the framework for projection, although they perform the major functions of testing hypotheses and of motivating their reformulation.

In contrast, comparative cost analysis does add a firm analytical scaffolding to many of the types of studies discussed in previous chapters. Although this point will be fully developed in the final chapters, we may briefly touch upon it here. For the derivation of population projections, industry-by-industry comparative cost studies provide a first approximation[60] to the regional pattern of new job opportunities. These opportunities in turn suggest a regional pattern for that large segment of population whose location is tied to economic opportunities. Moreover, comparison with natural rates of increase by region leads to first approximations of interregional migration of persons within this segment of population. The major fraction of changes in regional income can be anticipated directly from the regional pattern of new job opportunities and of projected industrial output. In like manner, implications for commodity flows, money flows, balance of payments, and cyclical sensitivity can be drawn with a fair amount of confidence, as we shall detail in Chapter 12.

Appendix A

Scaling and Latent Structure Techniques

In section B of this chapter, the promise of a systematic industry-by-industry comparative cost study for a region was indicated. Comcomitantly it was recognized that other important noneconomic factors—for example, political organization, community attitudes, cultural patterns, and business confidence—are at play in locational decisions. They are largely subjective in character and, accordingly have usually been treated in an intuitive manner. Recently, however, new quantitative techniques, which pertain particularly to attitude measurement and pattern identification, have been emerging in the fields of psychology and sociology. Although they have found little application in regional studies, their current promise for advancing the state of regional analysis is sufficient to justify a brief discussion in this Appendix.

1. Scaling techniques

To begin, suppose we consider a region, say New England. Suppose, too, it has been possible to pursue a systematic industry-by-industry comparative cost

[60] This approximation is, of course, subject to improvement through the use of techniques to be discussed in subsequent chapters. As noted, these techniques cut through certain important postulates which restrict comparative cost study.

study for this region. (Perhaps, too, it has been possible to supplement the comparative cost study with industrial complex and interregional input-output studies to be discussed in subsequent chapters.) After appropriate check on available resources—for example, port facilities, water of appropriate quality, road and rail transportation, labor skills, and power supplies—for each industry, a set of conclusions is reached with respect to the feasibility of locating a new plant in New England.

Such analysis is frequently insufficient, as the historical record testifies. New industrial development can be precluded not only by unfavorable cost conditions but also by unfavorable community attitudes and cultural patterns (even when cost conditions are very favorable). Further, not only must attitudes and patterns be generally favorable or at least neutral in the region (New England) as a whole; they must also be generally favorable, or at least not too hostile, in the particular community possessing the specific resources (such as a port facility) to which potential new industrial plants must be oriented. Put otherwise, there must be at least some spatial association of favorable community attitude and specific potential plant sites for industrial development to ensue even when other necessary conditions are met.

Measurement of attitude—favorable, neutral, unfavorable—is a problem which presents major difficulties, both conceptually and technically. One significant approach—the Guttman scaling or scalogram technique—attempts to identify a single scale along which effective measurement of attitude in a given situation can be attained.[61] Typically a set of questions are asked, each of which requires a "yes" or "no" answer. Ideally, these questions (items) are to be so phrased and arranged that a positive answer by a respondent to any given question implies or requires for consistency a positive answer by the respondent on all questions of lower position. As an obvious example, if the following questions are asked:

1. Do you weigh over 150 pounds?
2. Do you weigh over 125 pounds?
3. Do you weigh over 100 pounds?

a respondent who answers positively to question 1 must answer positively to questions 2 and 3 to be consistent.

As a second example, consider the general attitude (past and current) toward the Negro as reflected in institutional practices in four representative states, Virginia, Maryland, West Virginia, and Pennsylvania.[62] Such attitude may be scaled according to the presence or absence of certain characteristics: white primary (1); Jim Crow railways (2); and school segregation laws (3). In 1944, these four states ranked as indicated in Table A-1. Presence of characteristic (1) (corresponding to a positive response) thus implies presence of lower charac-

[61] For full discussion of this technique, the reader is referred to, among others, S. A. Stouffer et. al. [63], ch. 1–9; L. Guttman and E. A. Suchman [25]; M. W. Riley, J. W. Riley, J. Toby, et. al. [55]; L. Festinger and D. Katz [18], pp. 260–269, 525–528; M. J. Hagood and D. O. Price [29], pp. 144–152; G. Shapiro [60], pp. 619–621; N. E. Green [23], pp. 8–13; P. L. Lazarsfeld et. al. [49], pp. 216–257; J. A. Davis [10], pp. 371–380; and J. S. Coleman [7].

[62] The materials presented are selected, for pedagogical purposes only, from G. Shapiro [60], pp. 619–621.

teristics (2) and (3); and presence of characteristic (2) implies presence of lower characteristic (3).

Note that the Guttman scale is a cumulative-type scale. To repeat, it involves ranking respondents so that in a perfect scale each respondent will agree with or react positively to all items up to the point that represents his own attitude and disagree with all items beyond that point. (That is, respondents are arranged in order from those with the most positive attitude to those with the most negative attitude.) Or it involves ranking regions so that in a perfect scale each region possesses all characteristics up to the item representing its own position on the scale and lacks all characteristics beyond that item. But not only does the technique rank respondents or regions; it also arranges the items (questions, or characteristics) in order according to the relative position they represent along the scale of measurement. To be specific, in scaling the general attitude toward the Negro in the four states given, it is not apparent at the start which of the three characteristics (white primary, Jim Crow railways, or school segregation laws) represents the highest point on the scale measuring discrimination against the Negro. Once the data are compiled, it is evident that item (1) (white primary)

TABLE A-1. SCALOGRAM 1

	Presence of			Absence of		
State	(1)	(2)	(3)	(1)	(2)	(3)
Virginia	×	×	×			
Maryland		×	×	×		
West Virginia			×	×	×	
Pennsylvania				×	×	×

must represent the highest point along the assumed unidimensional scale. This statement follows since a state with three discriminatory characteristics must be ranked higher than a state with only two of these three characteristics. On a cumulative-type scale such can only be true if that characteristic not common to both states represents a higher point on the scale than any characteristic common to the two states. Thus, characteristic (1) must represent a higher point than either (2) or (3).

It is also evident that characteristic (2) must represent a higher point on the scale than (3). For, from the data given, one state (West Virginia) possesses characteristic (3) alone, whereas a second state (Maryland) possesses both characteristics (2) and (3). [Neither state possesses characteristic (1).] Since Maryland possesses both characteristics, it must be ranked higher than West Virginia. But this can only be if (2) represents a higher point on the cumulative, unidimensional scale.

Thus, on the assumption that characteristics (1), (2), and (3) measure the same attitude, we obtain a rank of states by intensity of discrimination.

Unfortunately, perfect unidimensional scales of the type illustrated are atypical. Usually, there is one or more deviations from the ideal pattern of Scalogram 1 (Table A-1). This situation obtains even though in the application of the technique the analyst tends to exclude items that cannot be arranged along a single scale. When there are many deviations from a perfect pattern, the analyst may search for scalable subareas or sectors; or, if such are not found, he may forgo the use of the technique.

These points can be more lucidly developed if we return to the discussion of the industrial development problem of New England. On the basis of past and current studies it would not appear fruitful to use the limited research resources available to scale the *general New England attitude* toward new industry. There is overwhelming evidence that a significant fraction of the area and its population favors industrial growth. A more relevant investigation would attempt to classify communities by attitude toward industrial development, and thereby to determine the extent to which advantageous potential plant sites exist in or around communities with favorable attitudes.

Although we could conduct a thorough and extensive attitude survey for each New England community, such a study would be costly and time consuming; and in view of the limitations of survey techniques, such an attack may not be justifiable. An alternative procedure, much less direct and perhaps inferior but much less costly and time consuming, is to (1) advance a reasonable hypothesis such as "a community's resistance to industrial development varies directly with its socio-economic status", and (2) attempt to classify New England communities according to their position along a unidimensional Guttman scale of socio-economic status. Such a scale would be based on existing sets of information, typically of a census variety such as per capita income, residential density, educational level, and home ownership.

As far as the authors are aware, no such scale study of socio-economic status of communities within a large region has been conducted. However, a scale study of census tracts, ranked by socio-economic status, has been undertaken for the metropolitan region of Birmingham, Alabama. Since a scale study for the communities of a region would likely parallel in the most important respects the completed study for Birmingham, we sketch it in order to illustrate the virtues and limitations of scaling techniques in general for regional analysis.

In the Birmingham study, 28 of the 58 census tracts in the city area were chosen as a representative sample. Five social-data items were selected for the development of a socio-economic status scale. These items are recorded in Table A-2. Note that these items are trichotomous rather than dichotomous. Three responses *A*(*negative*), *B*(*neutral*), and *C*(*positive*), are possible for each item. When the response on each of the five items is recorded for each of the 28 census tracts, 140 responses are obtained. Scale analysis of these 140 responses shows that the five items may be taken to represent a single-dimension scale applicable to these census tracts. The scalogram developed in this study is reproduced as Scalogram 2 (Table A-3).

In Scalogram 2 (Table A-3), the 28 census tracts are identified by number in column 2. The letters *C*, *B*, and *A* at the top of the table indicate positive, neutral, and negative responses, respectively. The numbers 1, 2, 3, 4 and 5 refer respectively to *income*, *crowding within dwellings*, *home ownership*, *social disorganization*, and *education*, as detailed in Table A-2. The response pattern for each census tract is noted, the particular arrangement presented being the one which the author of the study found to conform most closely to the ideal (perfect-scale) parallelogram depicted in Scalogram 1 (Table A-1). This particular arrangement, as already discussed, then determines the ranking of the census

tracts as well as the point on the scale represented by each type response to each item. For example, tract 21 ranks in the highest group (scale type I); and a positive response to item 1 (i.e., a position within the highest levels of median annual income) represents the highest point on the cumulative scale of socio-economic status.

Note that there are eleven responses which are not in place, that is, are deviations from the perfect parallelogram. The ratio of this number of deviations to 140, which is the total number of responses, measures in one sense the extent

TABLE A-2. ITEM AND CATEGORY DEFINITIONS FOR SCALE OF SOCIO-ECONOMIC STATUS

Item	Subject	*A*(Negative)	*B*(Neutral)	*C*(Positive)
1.	Median annual income of all employed persons	Lowest 7 ranks	Middle 14 ranks	Highest 7 ranks
2.	Prevalence of crowding within dwellings (1.01 or more persons per room)	Highest 7 ranks	Middle 14 ranks	Lowest 7 ranks
3.	Prevalence of home ownership (percentage of dwellings owner-occupied)	Lowest 7 ranks	Middle 14 ranks	Highest 7 ranks
4.	Prevalence of social disorganization (percentage of families involved in crime, delinquency, divorce, etc.)	Highest 7 ranks	Middle 14 ranks	Lowest 7 ranks
5.	Educational achievement (median years of school completed by persons 25 and over)	Lowest 7 ranks	Second lowest 7 ranks	Highest 14 ranks

Source: N. E. Green [23], p. 11.

to which the scalogram in and of itself fails to reproduce exactly the pattern of responses. Or, if this ratio is subtracted from unity, we obtain a coefficient which has been designated the *coefficient of reproducibility*, and which measures the extent to which the scalogram can reproduce the pattern of responses. Conventionally, a coefficient of reproducibility of at least 0.90 has been viewed as a necessary condition for a scalogram to have content. Other criteria should also be met, for example, criteria with respect to randomness of deviations. The reader is referred to the literature already cited for their discussion.

Note on Scalogram 2 that at the extreme left is a column indicating *scale type*. This column simply differentiates and ranks the different possible response

TABLE A-3. SCALOGRAM 2: SCALE OF SOCIO-ECONOMIC STATUS FOR TWENTY-EIGHT CENSUS TRACTS IN BIRMINGHAM, ALABAMA

(Coefficient of reproducibility = 0.92)

Item Number and Response Category

Scale Type	Tract Number	*C* 1	*C* 3	*C* 4	*C* 2	*C* 5	*B* 1	*B* 3	*B* 4	*B* 2	*B* 5	*A* 1	*A* 3	*A* 4	*A* 2	*A* 5	Scale Score
I	21	×	×	×	×	×											20
I	38	×	×	×	×	×											20
I	1	×	×	×	×	×											20
I	23	×	×		×	×			×								20
II	19		×	×	×	×	×										18
II	22		×		×	×	×		×								18
II	4		×		×	×	×		×								18
III	31	×		×	×	×		×									16
III	30			×	×	×	×	×									16
III	47	×		×	×	×		×									16
IV	3				×	×	×	×	×								14
IV	50				×	×	×	×	×								14
IV	40	×			×	×		×	×								14
V	34					×	×	×	×	×							12
VI	33						×	×	×		×				×		10
VI	42						×	×	×	×	×						10
VI	18						×	×	×	×	×						10
VI	9						×	×	×						×	×	10
VI	8						×	×	×	×	×						10
VI	5						×	×	×	×	×						10
VI	13						×	×	×	×	×						10
IX	27				×						×	×	×	×			4
IX	45									×		×	×	×		×	4
XI	26											×	×	×	×	×	0
XI	44											×	×	×	×	×	0
XI	43											×	×	×	×	×	0
XI	28											×	×	×	×	×	0
XI	46											×	×	×	×	×	0
Frequency		7	7	7	14	14	14	14	14	7	7	7	7	7	7	7	(140)
Errors		3	0	0	1	0	0	0	3	0	0	0	0	0	2	2	(11)

Source: N. E. Green [23], p. 12.

patterns for the perfect parallelogram which serves as the model for the scalogram; and tracts are assigned to a particular type as if they had no deviant responses. At the extreme right of Scalogram 2 is another column indicating the score of the corresponding scale type. It is based on arbitrary weights but can be useful for certain comparative purposes.

Now we return to the problem of industrial development in New England. Suppose a scalogram such as Scalogram 2 were developed for all communities above a specified size in New England. (For the moment we set aside the evaluation of such a scalogram.) According to our hypothesis, this scalogram would rank communities by degree of resistance to new industry. A check on this hypothesis would be obtained by ranking communities that have recently responded to the possibility of new industry. Those communities that have most successfully resisted the introduction of new industry should be among the highest (first) scale types in the socio-economic status continuum; those that have most actively encouraged new industry should be among the lowest scale types; and so forth. If there is not sufficient correspondence between rank by experience and rank on the Guttman scale, then clearly the hypothesis, the particular Guttman scale, or both are inappropriate.

If there is a reasonable correspondence between rank by experience and rank on the Guttman scale, the analyst can proceed to identify certain scale types (the lower ones) that are likely to be receptive to new industry as well as certain scale types (the higher ones) that are likely to be resistant to new industry.[63] He also would note the rank of *the communities that possess the specific plant sites potentially advantageous from a cost standpoint for new industry*. If most or all these communities are of the scale types classed as receptive to new industry, he might proceed to use his projections of industrial development based on comparative cost study with little if any modification. In contrast, if most or all these communities are of the scale types classed as resistant to new industry, he would be compelled to qualify seriously his projections of new industrial development. Such qualification would be particularly necessary if there were no indications that attitudes in these communities could be changed through educational efforts, economic pressures and other forces.

Finally, if a large number of these communities are of the scale types classed as neither receptive nor resistant, the analyst would need to qualify his comparative cost projections to some extent at least. The extent of qualification would be conditioned, say, by intensive survey analysis of perhaps a sample of these communities in order to appraise better the internal forces at play,[64] or by

[63] In effect, the analyst chooses "cutoff points" based on the historical record of community reactions. The two cutoff points implied in the text lead to a three-way classification of communities.

In other cases such cutoff points might be chosen arbitrarily at levels which would include a specific proportion of communities studied. This procedure is somewhat analogous to the choice of confidence limits in statistical research.

[64] It should be noted that when we construct a scalogram relating to an attitude where the response ranges from positive to negative, we may also construct an intensity function or curve. Such a function or curve reflects the strength of the attitude held by a respondent and is determined by asking a set of questions such as: "How strongly do you feel . . . ? Very Strongly, Fairly Strongly, Not So Strongly, or Not At All Strongly." A generally accepted hypothesis is that respondents in the extreme class types of a scalogram react much more intensively than respondents in the middle class types such

an estimation of promotion efforts which may be concentrated on these communities by business, governmental, and other groups, etc.[65]

The relevance of this application of scaling techniques in conjunction with comparative cost and other analysis obviously depends on the validity of the several techniques and hypotheses employed. Apropos the scaling technique, a number of major limitations should be recognized. First, the analyst may be forced at several steps in the scaling procedure to make arbitrary decisions on subjective grounds. The original choice of items thought relevant to the scale being sought depends frequently on the judgment of the researcher. Further, the subsequent arrangement of both items and respondents in the scalogram requires a subjective balancing of criteria, which criteria often have conflicting requirements. Consequently, on many occasions, when the coefficient of reproducibility is relatively low, different patterns may result from the same data as studied by different analysts.[66]

Beyond these considerations are certain basic problems of interpretation of results. Once a relevant scalogram has been constructed, how should we determine whether a pattern of deviation is random (quasi-scale) or nonrandom (i.e., indicative of the presence of other dimensions)? Moreover, what significance should be attached to a coefficient of reproducibility when the coefficient itself varies with the fineness of the steps between items? Still more, the scalogram procedure tends to eliminate from the study items (or characteristics) significant for the problem being attacked but yielding patterns too deviant (unique) to satisfy the scaling criteria.[67]

Despite these and other limitations—the reader is referred to the literature cited for their full discussion—the scaling technique has considerable potential and has in fact been extensively applied by psychologists and sociologists. As already noted, its value in the social sciences for measuring attitudes and identifying dimensions of social structure lies in its transformation of qualitative and noncomparable quantitative information into numerical rankings (ordinal values). Such rankings, moreover, permit the subsequent use of rank correlation, index construction, and other quantitative techniques.

that a U- or J-shaped intensity function results when intensity is plotted along the vertical axis and class types are plotted in order along the horizontal axis. The minimum point (designated the zero point) of such a function is held to divide a population into two sectors such as "for" and "against"; or "receptive" and "hostile." To the extent that these hypotheses are valid in a given situation, to that same extent they permit less intensive study of respondents in certain class types, in some cases those types clustered around the zero point as a cutoff point, in other cases those types clustered at the extreme. Generally speaking, such hypotheses allow economy in more extensive attitude investigations.

65 Obviously if this type of analysis is valid, it yields as a by-product vital information on the need for educational and similar efforts, if such are desirable, and the particular communities at which such efforts should be aimed.

66 For example, in Scalogram 2, census tracts 31 and 47 (both scale type III) have identical response patterns; yet they are separated in the ranking by tract 30 with a different response pattern. A similar situation exists in scale type IV.

67 For example, an item basic to the general attitudes toward industrial development in New England, say ethnic stock, may be eliminated from a scalogram on socio-economic status because of a unique pattern. This would reflect a limitation of both the scaling technique and the hypothesis relating attitude to socio-economic status.

More specifically, in regional analysis the scaling technique can be used for such a variety of purposes as (1) to estimate effectiveness of birth control and public health practices and other factors as they relate to key parameters of regional population projections; (2) to contrast attitudes of various groups of migrants, or of migrants versus nonmigrants, the better to estimate interregional and intraregional population movements; (3) to construct meaningful (unidimensional) categories of welfare and social accounts in nonmarket activities, especially in underdeveloped regions; (4) to determine more efficiently and accurately whether or not a regional population favors a particular resource development proposal or policy; (5) to judge whether governmental units in the several regions have sufficient authority and power to implement different regional programs (in a manner analogous to that suggested in the hypothetical case of community attitudes about industrial development in New England); (6) to identify groups of individuals who might be more receptive to soil conservation and similar resource development programs; and to plan an effective chronological sequence of administrative steps; and finally, (7) to uncover with more objectivity bonds (interrelations) among regions and subareas within any given region which are of an attitudinal-cultural nature.

2. LATENT STRUCTURE METHODS

Conceptually more satisfying, but operationally much more lean, are latent structure methods. These methods have conceptual appeal because they can successfully attack "nonscale" situations. As indicated earlier, nonscale situations are those involving response patterns (or patterns of characteristics) which are not satisfactorily scalable along a single dimension. However, these situations and their response patterns may be consistent with a meaningful underlying set of classes of respondents, where these classes are identifiable with respect to one or more dimensions. The latent structure framework aims at such identification and therefore represents a generalization of scaling techniques.[68]

Basic to latent structure methods is a reasoning process which starts with data obtained from relevant questionnaires and other empirical study and which may conveniently be termed *manifest data.* Given such data, a model is constructed which assumes the existence of a system of classes of respondents. Such classes are termed *latent* classes. Each of these classes is defined in terms of a set of probabilities. That is, for a given class each possible response pattern (such as those in Scalogram 2) is associated with a probability factor. More specifically, in Table A-4 are listed in column 1 all possible types of response patterns relating to four items (the sign + indictates a positive response and the sign − a negative response). Also, in column 3 are listed the probabilities associated with latent class I. Each figure in column 3 indicates the probability that the respondent who checks the corresponding response pattern belongs to class I. Thus, the figure of 0.995 at the top of column 3 indicates that anyone who checks off a + + + + response pattern has 995 chances out of a thousand of being in class I.

Behind the probabilities recorded for each latent class (e.g., those in columns

68 For full discussion of this technique, see especially S. A. Stouffer et al. [63], pp. 19–33, chs. 10, 11; P. F. Lazarsfeld et al. [49], pp. 349–387; P. F. Lazarsfeld [48], pp. 391–403; and L. Festinger and D. Katz [18], pp. 524–532.

3, 4, and 5 of Table A-4) is the basic mechanism which generates hypothetical frequencies for each response pattern listed in column 1. (This mechanism is too complex to develop in the brief scope of this Appendix.) This basic mechanism essentially determines another set of probabilities, namely the probabilities that any member of a *given class* will check off the several possible response patterns. Thus, if we wish to determine the total number of times a particular response pattern (say + − − +) will be found from (generated by) the operation of the

TABLE A-4. GENERATED DATA OF A HYPOTHETICAL LATENT STRUCTURE MODEL

(1)				(2)	(3)	(4)	(5)	(6)	(7)
Response Pattern Item				Observed Frequency	Probabilities by Latent Class			Sum of Cols. 3–5	Generated Frequency
1	2	3	4		I	II	III		
+	+	+	+	147	0.995	0.004	0.001	1.0	148.4
+	+	−	+	11	0.978	0.004	0.018	1.0	13.4
−	+	+	+	128	0.856	0.123	0.021	1.0	133.9
+	+	+	−	1	0.852	0.004	0.144	1.0	0.8
+	−	+	+	58	0.807	0.164	0.029	1.0	60.0
−	+	−	+	27	0.573	0.083	0.344	1.0	17.8
+	−	−	+	9	0.487	0.099	0.414	1.0	8.9
−	−	+	+	341	0.113	0.756	0.131	1.0	331.9
−	−	−	+	112	0.028	0.188	0.784	1.0	118.9
+	+	−	−	4	0.196	0.001	0.803	1.0	0.3
−	+	+	−	2	0.149	0.021	0.830	1.0	3.7
+	−	+	−	5	0.110	0.022	0.868	1.0	2.1
−	−	+	−	47	0.004	0.025	0.971	1.0	48.2
−	+	−	−	5	0.007	0.001	0.992	1.0	6.7
+	−	−	−	3	0.005	0.001	0.994	1.0	4.0
−	−	−	−	100	0.000	0.001	0.999	1.0	101.0
Totals in each class				1000	381.0	304.2	314.8	–	1000.0

Source: Data fictitious. Numerical figures identical with data in Table 11, Stouffer [63], p. 440.

model, we (1) take each class and multiply the number of respondents within it by the probability that its respondents will check off the particular response pattern (+ − − +), and (2) sum over all classes.

The last statement provides the basis for testing a model. An analyst simply compares for each response pattern the hypothetical frequency generated by the model (say column 7 of Table A-4) with the actual frequency as recorded in the manifest data (say column 2 of Table A-4). To the extent that there is a close correspondence, a correspondence closer than yielded by any other meaningful

model, he may infer that the classes of respondents postulated by the model exist and form a latent, underlying structure. Note that such a structure need not be ordered along any dimension.[69]

We have merely sketched the basic approach of the latent structure method.[70] Unfortunately, this method still requires extensive development before the major computational (as well as conceptual) problems associated with most of its potential applications can be overcome. (These computational problems stem from the need to solve complex systems of simultaneous equations in order to derive the parameters of the generating mechanism.) Therefore, we shall not go into further details in this Appendix. The reader is referred to the literature cited for full treatment. In the remaining paragraphs, however, we wish to indicate some directions for its potential use in regional analysis.

One direction of possible future use might, for example, be indicated by a study of industrial development in New England which has already been alluded to. Suppose the analyst finds it impossible to scale along the socio-economic dimension examined. At best, suppose he obtains a coefficient of reproducibility of 0.7, an unacceptable level. Such a finding does not preclude a more sophisticated analysis. It simply signifies that the response patterns are nonscalable. The analyst may still investigate a latent structure model to unearth the system of classes of respondents which may underlie the response patterns.

To illustrate let Table A-4 depict data on the response patterns (or patterns of characteristics) of communities of New England. Four basic questions (characteristics), represented by items 1, 2, 3 and 4 in column 1,[71] are considered relevant. A response indicating resistance (or the presence of an unfavorable characteristic, e.g. high median income) is indicated by a plus sign. A response indicating a receptive attitude (or the absence of an unfavorable characteristic) is indicated by a minus sign. All possible response patterns are listed; they number 16 (i.e., 2^n where n = the number of items). In column 2 are listed the number of communities having each pattern.

In line with this discussion, a model is constructed. According to Table A-4 this model is found consistent with three latent classes, I, II, and III. As noted before, the probability data generated by the model are presented in columns 3, 4, and 5. Again, each column refers to a particular latent class and shows for

[69] A system of classes is simply identified as a set of points in the positive quadrant of n-dimensional space, where the components of any point (each a probability) sum to unity, and where n is the number of possible response patterns. If the classes do lie along a unidimensional continuum, a latent structure analysis should give the same results as a scale analysis.

It should also be noted that the identification of a set of latent classes does not prove any causal hypothesis. However, it may be used to test hypotheses in the statistical sense, and in conjunction with other materials to suggest possibly significant causal hypotheses.

[70] It is also to be observed that latent structure analysis is somewhat analogous to factor analysis, which will be touched upon in Appendix B to this chapter. Factor analysis assumes that the variables with which an analyst deals are continuous and have normal joint distributions. Latent structure analysis utilizes items which are non-continuous (typically dichotomous or trichotomous) and does for such items a job similar to what factor analysis does for quantitative variables.

[71] These questions or characteristics may be similar to the items of Scalogram 2 or may relate to a host of other pertinent traits and features.

each response pattern the probability that a respondent (community) with that response pattern will be found in that latent class. And each row indicates for the relevant response pattern the probabilities that a respondent (community) with that response pattern will be found in the several latent classes. Therefore, the three probabilities along any row must add to unity as indicated in column 6. Finally, the frequency data generated by the model are recorded in column 7. Note that the generated data correspond well with the actual data; hence the model may be said to be a relatively good fit (the reader is referred to the literature for relevant tests of fit).

At this point it must be reiterated that the model does not furnish a basis for ordering classes. It merely tests the existence of classes. Therefore, it becomes necessary, at least in a number of instances, to introduce additional information in order to acquire further insight. In the particular problem of the industrial development of New England, the analyst may have worded his questions in such a manner, or have selected such characteristics, that the two response patterns, + + + + and − − − −, can be taken to represent only the two extreme positions along the single dimension of community resistance. Scrutiny of the generated data on probabilities (columns 3 to 5) does suggest that the model is consistent with the initial choice of questions or characteristics. The data do show that of all response patterns the extreme pattern + + + + has the greatest probability of being found in one class of communities. Accordingly this class, latent class I, can be considered as tending to be resistant to industrial development, especially when the probability that response pattern + + + + will be checked off by members of that class is comparatively high. (In the model behind Table A-4, this probability is approximately 0.37.) Simultaneously, the generated data do show that of all response patterns the extreme pattern − − − − has the greatest probability of being found in another class of communities. Accordingly, this class, latent class III, can be considered as tending to be receptive to industrial development, especially when the probability that response pattern − − − − will be checked off by members of that class is comparatively high. (This probability is approximately 0.32.)

From these assumptions and the derived partial order of response patterns, the analyst can proceed to certain conclusions. As before, he may determine cutoff points of significance. He may judge that, in addition to response pattern + + + +, the response patterns + + − +, − + + +, and + + + − are also suitably classified as resistant. For each respondent indicating one of these patterns the probability of belonging to latent class I (already designated as tending to be resistant) is 0.850 or more. Likewise, he may judge that, in addition to response pattern − − − −, the response patterns + − − −, − + − −, − − + −, and + − + − are also suitably classified as receptive. For each respondent indicating one of these patterns, the probability of belonging to latent class III (already designated as tending to be receptive) is 0.850 or more.[72]

The researcher may now examine the response patterns (characteristics) of those communities possessing specific potential plant sites based on resource availability, etc. If most of or all these communities have response patterns which he has classified as receptive, he may leave unqualified his conclusions based on comparative cost analysis. In contrast, if most of or all these communities have response patterns which he has classified as resistant, his comparative cost conclusions must be seriously qualified. Lastly, if most of or all these communities have response patterns which do not lend themselves to neat classification (as the central response patterns of Table A-4), he will need to qualify his conclusions. The extent of qualification will be influenced by the findings of other studies and by the anticipated effectiveness of any promotional or educational efforts by business or governmental units.[73]

[72] In determining cutoff points the analyst may pay attention to the distribution of probabilities among all classes, as well as the concentration in any one class.

We have illustrated one potential use of latent structure models. When the potential uses of scaling techniques which were listed at the end of the previous section involve situations that turn out to be non-scalable or not meaningfully depicted along a single dimension, the latent structure approach may be investigated. For example, migration phenomena might turn out to be non-analyzable in terms of a single dimension, and the analyst may judge it worthwhile to search for latent classes of migrants. If such classes of migrants are found, the motivating forces of each class can be studied in turn and can lead to firmer interregional and intraregional projection of flows of migrants.

There are other potential uses for the latent structure approach, such as to identify latent classes in a population from which representative individuals can be drawn to estimate, say, a community participation potential function which will be discussed in Chapter 11. But it would be premature at this time to detail any such potential application of latent structure approach; the conceptual and computation difficulties confronting the widespread use of this approach are both many and severe. Their enumeration and discussion are beyond the limited objective of this Appendix, which has been merely to sketch the promise of latent structure methods.

Appendix B

Factor Analysis, with Particular Reference to Regional Delineation

At a number of points in this book, the problem of selecting appropriate sets of regions for analysis has been alluded to. This problem is particularly acute in connection with the last section of this chapter concerned with coefficients of localization, specialization, redistribution, localization curves, shift ratios, etc. This problem is present in most regional investigations and is rarely fully resolved. This situation obtains not only because of different philosophical approaches and welfare values connected with regional studies, topics beyond the scope of this volume, but also because an analyst typically finds reasonable alternative interpretations of the same objective data for delineating regions. Nonetheless, certain techniques are available for objective treatment of the data so as to reduce the possibility of error (or inconsistency) in the areas where subjective judgment must be made. One of these techniques is factor analysis, a technique which has found some useful application in the delineation of meaningful regions, and which can profitably find greater use in other facets of regional analysis. As with scaling and latent structure techniques, we shall

[73] As with the scalogram, if this type of analysis has validity, it yields as a by-product valuable information on key points in particular communities to which educational and similar efforts might be directed, if such are desirable.

attempt only to sketch the basic elements of factor analysis and some of its potential applications. The reader is referred to the cited literature for fuller discussion.[74]

Like many other research methods, factor analysis is designed to develop a simple framework of factors whose interplay can adequately represent the inter-

TABLE B-1. HYPOTHETICAL INTERCORRELATIONS

Item	1	2	3	4	······	50	······	100
1. Per capita income ($)	*	0.36	0.42	0.24	······	−0.12	······	0.06
2. Industrial Employment (% of total)	0.36	*	0.42	0.24	······	−0.12	······	0.06
3. Years of schooling, average	0.42	0.42	*	0.28	······	−0.14	······	0.07
4. Divorce rate	0.24	0 24	0.28	*	······	−0.08	······	0.04
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
50. Miles of highway per capita	−0.12	−0.12	−0.14	−0.08	······	*	······	−0.02
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.
100. Household accidents per capita	0.06	0.06	0.07	0.04	······	−0.02	······	*

* Data for intercorrelation of any one characteristic are not included since they are meaningless.

action of the complex set of forces in actuality. It has much in common with scalogram and latent structure analysis in that it attempts to combine or reduce variables which are linked to each other into indexes describing particular basic dimensions, or reflecting basic structural features of the total situation being studied; no dependent variable as such need be specified. It has less in common

[74] Among others, the following are useful general references: L. L. Thurstone [66]; K. J. Holzinger and H. H. Harman [32]; R. B. Cattell [6]; S. Stouffer et al. [63]; L. Festinger and D. Katz [18], especially pp. 274–278; B. Fruchter [21]; C. J. Adcock [1]; and M. J. Hagood and D. O. Price [29], ch. 26.

with regression and variance analysis for it does not attempt to explain statistically variation in a dependent variable by variation in a set of key independent variables, the discarded independent variables being judged as relatively insignificant. Rather it retains the many variables relevant in a study by attempting to account for their behavior in terms of relatively few basic dimensions.

To motivate the discussion, let us consider a simple case. Let the United States comprise the area corresponding to a system of regions. The problem is to divide the United States into a specific set of meaningful regions. For the problem to be studied, let us assume that an abundance of data is available, but only by state units. Hence for operational purposes, each region must be composed of whole states.

As a first step, a number of characteristics are to be selected, where the variation in each characteristic is hypothesized to reflect significantly the differentiation among underlying regions. Suppose that 100 characteristics are identified, as listed in the left-hand tab of Table B-1. Suppose, too, the intercorrelations of the 48 state scores on each pair of these characteristics are computed and recorded in the same table. If these intercorrelations take the specific "pure" form shown in the table, certain generalizations can be readily made. Since we may hypothesize at the start a single basic factor—namely a single basic set of meaningful regions—we may schematically represent this factor by the circle in Figure B-1. This circle cuts across a series of rectangles, each rectangle representing a particular characteristic. The amount of overlap with each rectangle indicates the extent to which the general factor accounts for (statistically explains) the variation among states in the corresponding characteristic. By each area of overlap is placed a decimal figure, which is customarily designated a factor loading. Squaring this factor loading and multiplying by 100 yields the per cent of the area of the corresponding rectangle which overlaps the circle; that is, for the characteristic represented by that rectangle it yields the per cent of the variation among states which is associated with regional differentiation. (In this respect as well as many others, the factor loading behaves like a correlation coefficient.) For example, the factor loading on characteristic 1 (per capita income) is 0.6. Squaring and multiplying by 100, we obtain the percentage figure of 36; hence 36 per cent of the variation among states in per capita income is to be explained, according to our hypothesis, by the basic factor of regions. Similarly, this basic factor statistically explains 36 per cent of the variation among states in characteristic 2 (industrial employment as a per cent of total employment), 49 per cent of variation in characteristic 3 (average number of years of schooling), etc.[75]

Given the hypothetical data of Figure B-1, and on the very important assumption that there is no other factor which relates any pair of characteristics, we may calculate expected intercorrelation between any two characteristics. (For the moment, the dashed ellipse coursing through rectangles 3 and 4 is to be ignored.) We simply compute the product of the two decimal factor loadings. For example, the expected intercorrelation between characteristics 2 and 3 is the product of 0.6 and 0.7, which product (0.42) is found in both the cell of row 2 and column 3 and the cell of row 3 and column 2 in Table B-1.

Thus the relationships depicted in Figure B-1 (when the dashed ellipse is nonexistent) afford in a pure statistical sense an explanation of all the intercorrelations which have been recorded in Table B-1. Or conversely the set of intercorrelations

[75] Note that the basic factor statistically explains 4 per cent of the variation in characteristic 50 which varies *inversely* relative to the other characteristics.

of Table B-1 implies one and only one basic factor and the specific factor loadings of Figure B-1. This strong implication results because the intercorrelations were deliberately constructed to be pure or, in other words, to be 100 per cent predictable by a single basic factor. That is, all columns (excluding

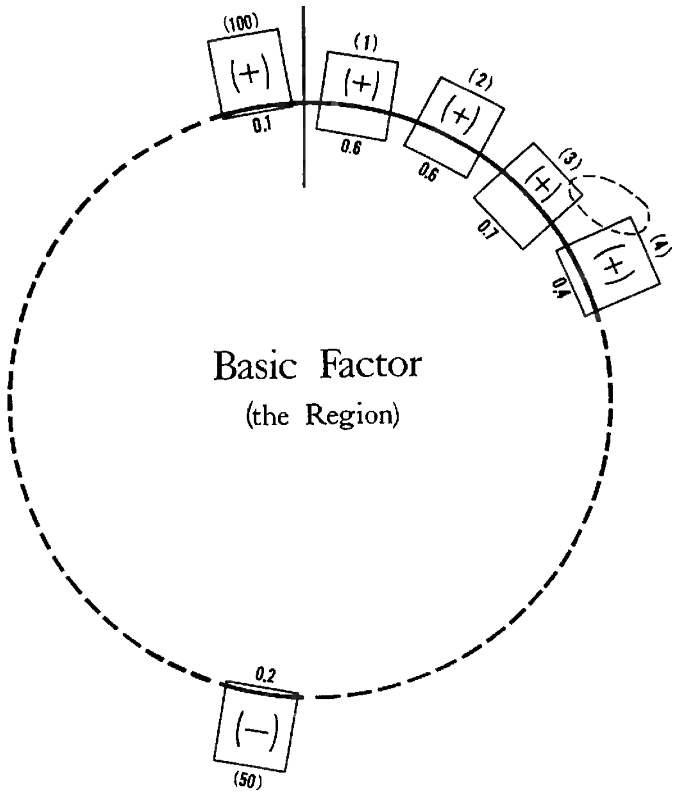


Figure B-1. Hypothetical factor-loading diagram.

items along the principal diagonal) are proportional to each other.[76] (For example, the 100th column is one-sixth of the first column.) Therefore the intercorrelations can reflect the play of only one general factor, even though this factor has a different impact on each characteristic. However, exactly what this

[76] In matrix algebra parlance, the rank of the intercorrelation matrix is one.

factor is, or how the result is to be interpreted, must rest solely on the analyst's conceptual and theoretical frameworks. (Recall that the choice of the characteristics is also largely determined by these frameworks.) In this simple example, we have hypothesized a system of regions which can be differentiated and whose diverse influences pervade the national system. We may then say that factor analysis has shown that the data are consistent with this hypothesis, and that only one basic factor exists. We interpret this factor as signifying that regions are present and do have differential impact. Yet, many other interpretations are possible, as the reader may discover from his own explorations with diverse hypotheses.

Once a single basic factor is found and interpreted as verifying the presence of a set of regions, the specific arrangement of states into regions can proceed via the construction of an index number and determination of cutoff points. Procedure for this step will be presented later when we report on some of Hagood's work.

Unfortunately, pure cases of this sort are rarely, if ever, found. A typical situation involves the interplay of a host of nonrandom factors as well as chance, as the reader by now fully appreciates. The intercorrelation data are then exceedingly complex. Yet this situation does not preclude a fruitful search for a single general factor in cases where strong theoretical support for such an hypothesis can be mustered. For example, in the delineation of single-purpose regions or of regions to be distinguished by a relatively narrow set of related characteristics, a single-factor approach may have considerable justification. However, when multiple-purpose regions are sought, or regions classifiable by the totality of characteristics, the single-factor approach, when intercorrelations are highly impure, is not readily accepted. Some of the pertinent reservations can be illustrated with reference to the pioneering and stimulating effort of Odum, Hagood, and others to demarcate a best set of major regions for the United States.

Hagood, who has perhaps developed most thoughtfully and carefully the single-factor approach in the delineation of regions, begins, as the factor analyst must, with a selection of relevant characteristics of states from which regions are to be fashioned. (It must be borne in mind that the limited availability of data on both state and other areal units may at the outset significantly restrict the extent to which valid results and tests are achievable.) Seeking a set of regions to be distinguished primarily by agricultural and demographic characteristics, she selects 104 characteristics, 52 agricultural and 52 demographic. The 52 agricultural characteristics are grouped into 6 classes, and the demographic into 8 classes. For each of the resulting 14 classes, a single-factor analysis is pursued. For example, one of the 6 agricultural classes is designated land use and covers the items listed in Table B-2.[77] The state values for each pair of these characteristics are correlated; the coefficients are listed in bold type in Table B-3. Application of standard computational procedures for single-factor analysis yields the factor loadings in column 1 of Table B-2. Thus the square of the factor loading of 0.540 at the head of the column indicates that the single factor (however interpreted) accounts for 29 per cent of the variation among states in the per cent farmland is of all land.

[77] The other agricultural classes relate to crops, livestock, tenure, farm values and farm finance.

TABLE B-2. SINGLE-FACTOR LOADINGS FOR LAND USE INDEX

Item	Factor Loading
	(1)
1. Per cent farmland is of all land	0.540
2. Per cent cropland (harvested and failure) is of all farmland	0.789
3. Per cent woodland is of all farmland not used for crops	0.479
4. Mean size of farm (acres)	−0.760

Source: Computed from data in M. J. Hagood [27].

As already noted, once factor loadings are obtained, "expected" intercorrelations when no other common factor is at play can be computed by multiplying the relevant pair of loadings. These expected intercorrelations are recorded in parentheses in Table B-3. Comparison of the "expected" with the "actual" data yields one test of the adequacy of the single-factor hypothesis. The considerable discrepancies of the data of Table B-3 do suggest the operation of one or more additional factors which are common to two or more of the four items (characteristics).

However, if the investigator does judge that the empirical results do not invalidate his single-factor hypothesis, he can proceed to construct an index, as Hagood does. For each state the index value is computed from the following equation:

$$I^J = 0.540Z_1^J + 0.789Z_2^J + 0.479Z_3^J - 0.760Z_4^J$$

where Z_1^J, Z_2^J, Z_3^J, and Z_4^J are the ratings of state J ($J = 1, \ldots, 48$) on each of the

TABLE B-3. INTERCORRELATIONS: ACTUAL AND IDEAL

Item	1	2	3	4
1		**0.624** (0.426)	**−0.091** (0.259)	**−0.189** (−0.410)
2	**0.624** (0.426)		**0.169** (0.378)	**−0.590** (−0.600)
3	**−0.091** (0.259)	**0.169** (0.378)		**−0.597** (−0.364)
4	**−0.189** (−0.410)	**−0.590** (−0.600)	**−0.597** (−0.364)	

Source: Derived from M. J. Hagood [27].

four characteristics of Table B-2. Each rating is in standard form (i.e., $Z_i^J = (X_i^J - M_i)/\sigma_i$ where $i = 1, \ldots, 4$; and where X_i^J is the actual value for characteristic i in state J, M_i is the mean value of the characteristic i over all states, and σ_i is the standard deviation of the state values). The coefficients which respectively multiply $Z_1^J, \ldots, Z_4^J$ are the factor loadings of Table B-2.

Once a set of index values for states is obtained, these values can be examined for cutoff points. If cutoff points can be located to set apart groups of states which are contiguous and on other counts can be expected to be homogeneous with respect to the phenomena being studied, it may be said that a single-factor analysis has helped in the objective determination of regions. For example, Map B-1 reproduces in a modified form the land use index map developed by Hagood. There is at least some indication of regional structure in this map.

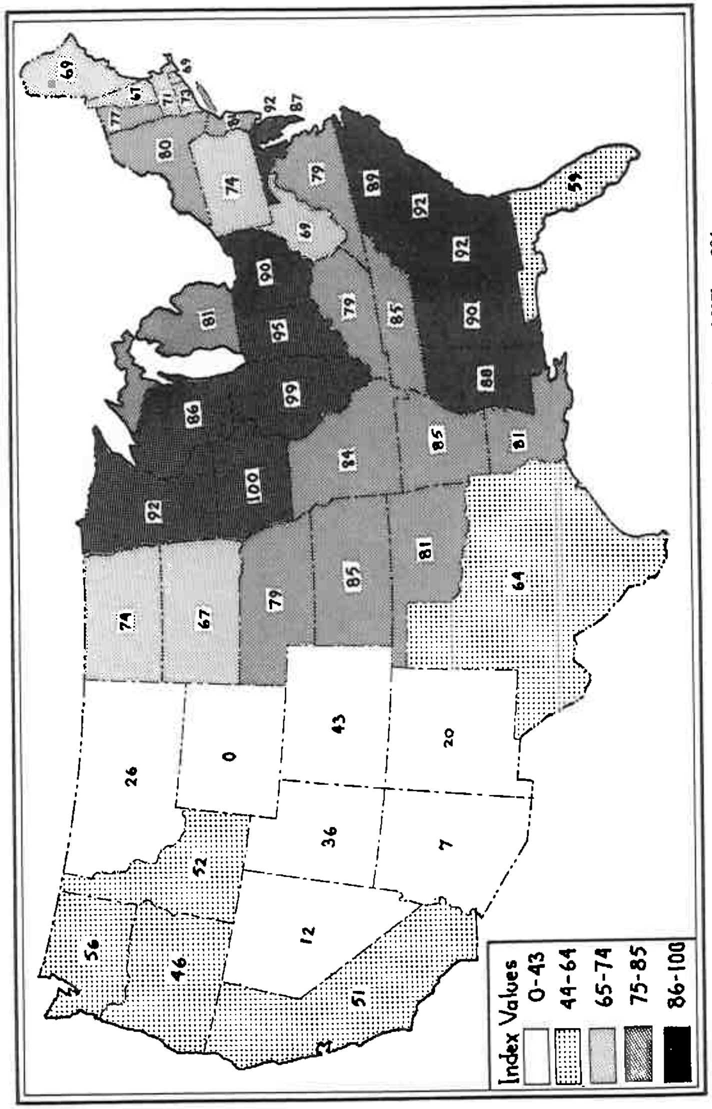
Hagood proceeds further. Once she obtains a land use index and an index for each of the other thirteen classes, she condenses these indices into two major group indices. One major group index is obtained from a single-factor analysis for the six indices representing the six agricultural classes. The second major group index is derived from another single-factor analysis for the eight indices representing the eight agricultural classes. This operation can lead to the determination of two additional sets of regions, a set of agricultural regions and a set of demographic regions.

Finally, Hagood performs a single-factor analysis on all fourteen indices, representing the fourteen classes of characteristics. Before actually combining states to form regions on the basis of significant cutoff points on her derived composite agriculture-population index, she takes another essential step. States may have the same index value because they have identical patterns (profiles) with respect to the 104 characteristics initially selected for study. They also may have the same value because differences in their patterns (which differences may be sharp) have compensatory effects on the index scale when weighted by factor loadings. (For example, in constructing an index based on the six classes of agricultural characteristics, Hagood finds that the states of Arizona and Iowa, which are highly dissimilar with respect to agricultural profiles, score 59 and 62 respectively.) Therefore another criterion for combining two or more states into one region is that they have similar patterns, to be evidenced for any pair of these states by positive and fairly high correlations between the values for the 104 selected characteristics.[78]

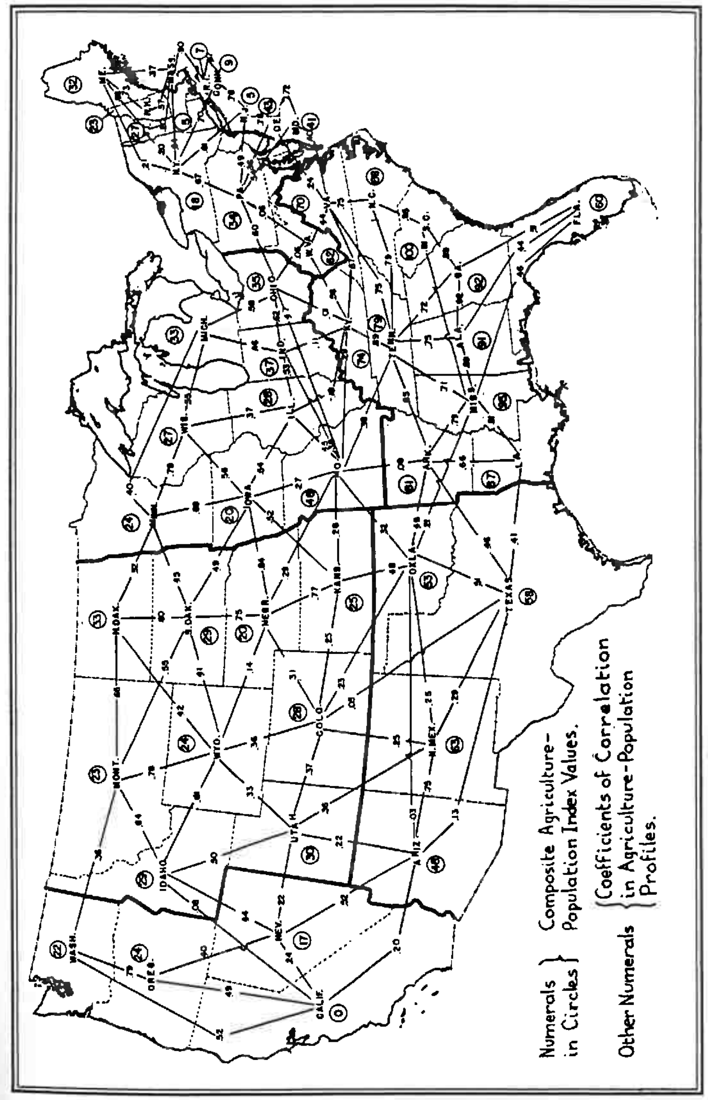
With the resulting index values and correlation coefficients, it becomes possible to organize states into regions, although a number of subjective elements still remain, as Hagood well recognizes. One set of possible regions is indicated in Map B-2 which is reproduced from Hagood's study. On this map the composite agriculture-population index value for each state is in bold type and is encircled. The correlation coefficients between the profiles (based on 104 characteristics) for selected pairs of adjacent and nearby states are indicated in light type.

Close scrutiny of Map B-2 does reveal that a number of sets of regions are possible. On the other hand, it does uncover certain clusters of states which, in a sense, tend to form the nuclei of regions (e.g., Mississippi, Alabama, Georgia,

[78] States within a region should also be contiguous. In another connection, Hagood has introduced latitudinal and longitudinal positions as relevant characteristics in factor analysis so that index values will to some extent reflect contiguity. See M. J. Hagood [26].



Map B-1. Land use index, 48 states, 1940. Source: M. J. Hagood [27], p. 291.



Map B-2. Composite agriculture-population index values and coefficients of correlation in agriculture-population profiles, 48 states, 1940. Source: M. J. Hagood [27], p. 296.

and South Carolina; Maine, New Hampshire, and Vermont; and Idaho, Montana, and Wyoming). Since state political units are poor units for the construction of regions, and since in a highly urbanized society the peripheral areas of regions are transitional, that is, are the zones where phenomena reflect mixed orientation, it is not unexpected that several alternative sets of regions are possible. In particular, it is not unexpected that alternative sets of regions may be formed by different assignments of such states as Missouri and West Virginia which contain or comprise transition areas.

In brief, it may be concluded that single-factor analysis of the type depicted can be of considerable value as an objective tool to complement theory and other analysis in the delineation of regions.[79] Yet we must recognize certain severe limitations in the use of this tool. As already noted, the single-factor approach finds its greatest validity when in fact all intercorrelations of characteristics are pure, that is, 100 per cent predictable from the single-factor loadings. But when these intercorrelations are not pure, difficult problems of interpretation arise.

We can illustrate this last point with reference to Table B-3. There the actual intercorrelations (boldface type) contrast rather sharply in a number of the cells with the expected intercorrelations (lightface type) based on the single-factor loadings of Table B-2. Hence many analysts would conclude that other common factors are at play. Such factors may account for part of or, at the extreme, all the discrepancy between the actual and expected intercorrelations for one or more pairs of characteristics. For example, in Figure B-1 the dashed ellipse coursing through rectangles 3 and 4 may be taken to represent a second factor common to characteristics 3 and 4, and therefore to explain an amount of correlation (if positive) over and above that explained by the first basic factor (the large circle). If other common factors are at play, then how interpret the findings from single-factor analysis?

One possible interpretation, which is implied by the work of Odum and Hagood and which would seem reasonable if the discrepancies are smallish, is that the findings from single-factor analysis are the really significant findings. An analyst might argue that although there are many factors in operation, a single one is dominant, and that in view of other limitations of the study, the presence of any secondary factors may be safely ignored. This line of reasoning might be particularly valid for the time being if the most relevant theoretical construct were oriented to a single cause-effect relationship.

But other lines of reasoning may also be plausible, in particular when the discrepancies are not small. An analyst's hypothesis may suggest two basic factors at play, or three, or four, etc. Factor analysis would then be required to unearth these several factors, for which task computation procedures have been developed. But, aside from the thorny theoretical problem of determining how many basic factors are at play, there is the major difficulty of interpreting results of multiple-factor analysis, even when the analytical framework unambiguously denotes the number of relevant factors.

To illustrate, set up a simple two-factor hypothesis relating to only four relevant characteristics, say any four listed in Table B-1. Let the actual intercorrelations be as recorded in Table B-4. Factor analysis operations, by one of the several standard techniques, yield loadings for factors I and II as indicated in Table B-5. Now, if we square 0.6, the first factor loading for characteristic 1, and add the result to the square of 0.4, the second factor loading for characteristic 1, we obtain 0.52. This number multiplied by 100 yields the per cent of variation in characteristic 1 which can be explained by the two factors. Also note that the intercorrelations of Table B-4 are pure, that is, fully explainable by the two factors. For example, on the basis of factor I alone we can expect intercorrelation between characteristics 2 and 3 of $0.6 \times 0.7 = 0.42$. On the basis of factor II alone we can expect intercorrelation between the same two characteristics of $0.6 \times -0.3 = -0.18$. Summing over both factors, we can expect intercorrelation of $0.42 - 0.18 = 0.24$ which is identical with the "actual" intercorrelation recorded in both the cell of row 3, column 2, and the cell of row 2, column 3.

TABLE B-4. HYPOTHETICAL INTERCORRELATIONS

Item	1	2	3	4
1		0.60	0.30	0.04
2	0.60		0.24	−0.06
3	0.30	0.24		0.43
4	0.04	−0.06	0.43	

TABLE B-5. HYPOTHETICAL FACTOR LOADINGS

Characteristic (Variable)	Factor I	Factor II
1	0.6	0.4
2	0.6	0.6
3	0.7	−0.3
4	0.4	−0.5

Despite the appearance of objectivity in the determination of the loadings for factors I and II, such objectivity does not in fact exist. For it can easily be demonstrated that many other sets of loadings for only two factors will explain just as well the intercorrelations of Table B-4. In fact, there is an infinity of such sets. The loadings of Table B-6 illustrate one of this infinity, as the reader

[79] A somewhat similar tool for the delineation of meaningful spatial units, but one less sophisticated from a statistical standpoint, is developed in E. Shevky and W. Bell [61]. For relevant evaluation see A. H. Hawley and O. D. Duncan [31]; and M. D. Van Arsdol, Jr., S. F. Camilleri, and C. F. Schmid [67].

may verify.[80] It thus becomes clear that, even for an hypothesis based on the operation of two and only two basic factors and even when all intercorrelations are fully explainable, multiple interpretations are possible. These different interpretations revolve around the diverse sets of factor loadings which can be considered pertinent to the problem studied.

From this discussion it is also clear that two-factor hypotheses where intercorrelations are only partially explainable require still more subjective judgment and intuition on the part of the analyst. Likewise with hypotheses involving three or more factors, whether or not intercorrelations are fully explainable in the statistical sense.

Although we have only touched on some of the basic features of factor analysis—the reader is referred to the cited literature for full discussion—some of its chief virtues and limitations are clear. It does offer a fruitful approach to condensing voluminous sets of data into relatively few useful indices or dimensions. As already illustrated, it can be an effective tool for delineating regions within a system which has firm theoretical foundations. It is a useful tool in constructing level-of-living indices[81] which along with correlation analysis of

TABLE B-6. ALTERNATIVE FACTOR LOADINGS

Characteristic (Variable)	Factor I	Factor II
1	0.08	0.72
2	−0.07	0.85
3	0.68	0.34
4	0.64	−0.02

profiles may aid comparative regional analysis. In certain connections it can furnish a useful basis for stratification in sampling.[82] As many statistical tools, it can serve as a partial test of an hypothesis or reflect on the adequacy of the characteristics initially selected as relevant, etc.[83]

Moreover, whether or not factor analysis is employed to test an hypothesis, it may suggest workable typologies and classification schemes and fruitful models or conceptual frameworks on the interrelations of variables. As one illustration, Price has employed factor analysis in searching for fundamental dimensions of metropolitan centers. He is able to explain statistically the intercorrelations in the variation of 15 characteristics among 93 cities in terms of 4 factors. The

[80] For example, multiplying −0.07 by 0.68 (factor I loadings on characteristics 2 and 3) yields −0.0476 which when added to 0.2890 (the product of 0.85 and 0.34, the respective factor II loadings) gives 0.2414, which is within rounding error of the value of 0.24 of Table B-4.

In factor analysis parlance, the loadings of Table B-6 are obtained by rotating the reference axes clockwise 50 degrees.

[81] For example, see A. L. Ferriss [17].

[82] For example, see M. J. Hagood and E. H. Bernert [28].

[83] For example, see M. D. Van Arsdol, S. F. Camilleri, and C. F. Schmid [67].

factor loadings obtained combined with his accumulated knowledge and judgment on forces at play in metropolitan centers leads him to identify tentatively these factors as (1) degree of maturity of city; (2) the extent to which a city is a service center; (3) the level of living within a city; and (4) the per capita trade volume of a city.[84] Finally, factor analysis (perhaps in conjunction with scalogram and latent structure methods) may be helpful in research on attitudes, political participation, and related topics which have an important bearing on resource development and planning for a region or system of regions. It may make possible the narrowing down of the range of alternative interpretations of complex sets of data, often incomplete, where such data might relate to voting patterns, use of government facilities (e.g., health centers, libraries, adult education programs, extension services), contacts among different social groups and institutions, etc. It may even suggest alternative hypotheses not apparent from scrutiny of raw data.

In considering the virtues of factor analysis, the analyst must also bear in mind the extent to which factor analysis cannot eliminate his responsibility for sound reasoning and judgment, and in many cases cannot eliminate the need to resort to arbitrary procedures. Briefly put, factor analysis is not nearly as objective as appears to the unsophisticated analyst. At the very start, the choice of relevant characteristics must depend on an investigator's intuition and previous knowledge, as well as the availability of data. Next, the number of factors deemed appropriate involves a judgment factor despite certain procedures designed to furnish objective criteria. Most important of all, many alternative sets of factor loadings are possible for a given set of data on intercorrelations, the particular one chosen and its interpretation (if any can be put forth) being largely determined by the theoretical or conceptual construct deemed most significant. In addition, there are a number of shortcomings both conceptual and technical, such as those relating to the assumptions involved in the initial correlation procedures, the additive or multiplicative nature of factors, and measures of error variance. The discussion of these shortcomings is beyond the scope of this Appendix.

References

1. Adcock, C. J., *Factorial Analysis for Non-Mathematicians*, Melbourne University Press, Melbourne, Australia, 1954.
2. Airov, Joseph, "Location Factors in Synthetic Fiber Production," *Papers and Proceedings of the Regional Science Association*, Vol. 2 (1956).
3. ———, *The Location of the Synthetic Fiber Industry: A Study in Regional Analysis*, John Wiley, New York, 1959.
4. Alexander, J. W., "Location of Manufacturing: Methods of Measurement," *Annals of the Association of American Geographers*, Vol. 48 (March 1958).
5. Bachi, Roberto, *Statistical Analysis of Geographical Series*, Kaplan School, Hebrew University and Israel Central Bureau of Statistics, Jerusalem, 1957.
6. Cattell, R. B., *Factor Analysis: An Introduction and Manual for the Psychologist and Social Scientist*, Harper, New York, 1952.
7. Coleman, J. S., "Multi-dimensional Scale Analysis," *American Journal of Sociology*, Vol. 63 (Nov. 1957).

[84] D. O. Price [54], pp. 449–455.

8. Creamer, Daniel, "Shifts of Manufacturing Industries," in *Industrial Location and National Resources*, U. S. National Resources Planning Board, Washington, D.C., 1943.
9. Cumberland, John, *The Locational Structure of the East Coast Steel Industry with Emphasis on the Feasibility of an Integrated New England Steel Mill*, doctoral dissertation, Harvard University, Cambridge, Massachusetts, 1951.
10. Davis, J. A., "On Criteria for Scale Relationships," *American Journal of Sociology*, Vol. 63 (Jan. 1958).
11. Dean, William H., Jr., *The Theory of the Geographic Location of Economic Activities*, doctoral dissertation, Harvard University, selections published by Edward Brothers, Inc., Ann Arbor, Michigan, 1938.
12. Duncan, Otis D., "Urbanization and Retail Specialization," *Social Forces*, Vol. 30 (March 1952).
13. ———, R. P. Cuzzort, and Beverly Duncan, *Statistical Geography*, The Free Press, Glencoe, Illinois, 1960.
14. ———, and Beverly Duncan, "A Methodological Analysis of Segregation Indexes," *American Sociological Review*, Vol. 20 (April 1955).
15. ———, and ———, "Residential Distribution and Occupational Stratification," *American Journal of Sociology*, Vol. 60 (March 1955).
16. Dunn, Edgar S., *The Location of Agricultural Production*, University of Florida Press, Gainesville, Florida, 1954.
17. Ferris A. L., "Rural-Farm Level of Living Indexes for Two Southeastern States," *Social Forces*, Vol. 26 (May 1948).
18. Festinger, Leon, and D. Katz, *Research Methods in the Behavioral Sciences*, Dryden Press, New York, 1953.
19. Florence, P. Sargant, *Investment, Location, and Size of Plant*, University Press, Cambridge, England, 1948.
20. ———, W. G. Fritz, and R. C. Gilles, "Measures of Industrial Distribution," in *Industrial Location and National Resources*, U. S. National Resources Planning Board, Washington, D.C., 1943, ch. 5.
21. Fruchter, B., *Introduction to Factor Analysis*, Van Nostrand, New York, 1954.
22. Fuchs, Victor R., "Changes in the Location of U. S. Manufacturing Since 1929," *Journal of Regional Science*, Vol. 1 (Spring 1959).
23. Green, N. E., "Scale Analysis of Urban Structures," *American Sociological Review*, Vol. 21 (Feb. 1956).
24. Greenhut, Melvin L., *Plant Location in Theory and Practice*, University of North Carolina Press, Chapel Hill, North Carolina, 1956.
25. Guttman, L., and E. A. Suchman, "Intensity and Zero Point for Attitude Analysis," *American Sociological Review*, Vol. 12 (Feb. 1947).
26. Hagood, M. J., "An Examination of the Use of Factor Analysis in the Problem of Subregional Delineation," *Rural Sociology*, Vol. 6 (Sept. 1941).
27. ———, "Statistical Methods for Delineation of Regions Applied to Data on Agriculture and Population," *Social Forces*, Vol. 21 (March 1943).
28. ———, and E. H. Bernert, "Component Indexes as a Basis for Stratification in Sampling," *Journal of the American Statistical Association*, Vol. 40 (Sept. 1945).
29. ———, and Daniel O. Price, *Statistics for Sociologists*, Henry Holt, New York, 1952.
30. Hauser, Phillip, M., Otis D. Duncan, and Beverly Duncan, *Methods of Urban Analysis: A Summary Report*, Research Report, U. S. Air Force Personnel and Training Research Center, San Antonio, Texas, 1956.

31. Hawley, A. H., and O. D. Duncan, "Social Area Analysis: A Critical Appraisal," *Land Economics*, Vol. 33 (Nov. 1957).
32. Holzinger, K. J., and H. H. Harman, *Factor Anlaysis: A Synthesis of Factorial Methods*, University of Chicago Press, Chicago, 1941.
33. Hoover, Edgar M., *Location of Economic Activity*, McGraw-Hill, New York 1948.
34. ———, *Location Theory and the Shoe and Leather Industries*, Harvard University Press, Cambridge, Massachusetts, 1937.
35. ———, "The Measurement of Industrial Localization," *Review of Economics and Statistics*, Vol. 18 (Nov. 1936).
36. ———, "Redistribution of Population, 1850–1940," *The Journal of Economic History*, Vol. 1 (Nov. 1941).
37. ———, and Joseph L. Fisher, "Research in Regional Economic Growth," *Problems in the Study of Economic Growth*, Universities-National Bureau Committee on Economic Research, National Bureau of Economic Research, New York, 1949.
38. Isard, Walter, *Location and Space-Economy*, John Wiley, New York, 1956.
39. ———, "Some Locational Factors in the Iron and Steel Industry Since the Early Nineteenth Century," *Journal of Political Economy*, Vol. 56 (June 1948).
40. ———, and William M. Capron, "The Future Locational Pattern of Iron and Steel Production in the United States," *Journal of Political Economy*, Vol. 57 (April 1949).
41. ———, and John Cumberland, "New England as a Possible Location for an Integrated Iron and Steel Works," *Economic Geography*, Vol. 26 (Oct. 1950).
42. ———, and Eugene W. Schooler, *Location Factors in the Petrochemical Industry*, Office of Technical Services, U. S. Department of Commerce, Washington, D.C., 1955.
43. ———, and Vincent H. Whitney, *Atomic Power, an Economic and Social Analysis*, McGraw-Hill, New York, 1952.
44. Krutilla, John V., *The Structure of Costs and Regional Advantage in Primary Aluminum Production*, doctoral dissertation, Harvard University, Cambridge, Massachusetts, 1952.
45. Kuenne, Robert E., *The Use of Input-Output Techniques for the Estimation of Employment in the Delaware Valley*, doctoral dissertation, Harvard University, 1953.
46. Launhardt, Wilhelm, "Die Bestimmung des Zweckmässigsten Standortes einer gewerblichen Anlage," *Zeitschrift des Vereins deutschen Ingenieure*, Vol. 26, No. 3 (1882).
47. ———, *Mathematische Begründung der Volkswirtschaftslehre*, Leipzig, 1885.
48. Lazarsfeld, Paul F., "Recent Developments in Latent Structure Analysis," *Sociometry*, Vol. 18 (1955).
49. ——— et al., *Mathematical Thinking in the Social Sciences*, Free Press, Glencoe, Illinois, 1954.
50. Lindsay, John Robert, *The Location of Oil Refining in the United States*, doctoral dissertation, Harvard University, Cambridge, Massachusetts, 1954.
51. ———, "Regional Advantage in Oil Refining," *Papers and Proceedings of the Regional Science Association*, Vol. 2 (1956).
52. Lösch, August, *The Economics of Location*, Yale University Press, New Haven, Connecticut, 1954.
53. Palander, Tord, *Beitrage zur Standortstheorie*, Almqvist & Wiksells Boktryckeri-A.-B., Uppsala, Sweden, 1935.

54. Price, Daniel O., "Factor Analysis in the Study of Metropolitan Centers," *Social Forces*, Vol 20 (May 1942).
55. Riley, M. W., J. W. Riley, J. Toby, et al., *Sociological Studies in Scale Analysis*, Rutgers University Press, New Brunswick, New Jersey, 1954.
56. Rodgers, Allan, "Some Aspects of Industrial Diversification in the United States," *Papers and Proceedings of the Regional Science Association*, Vol. 1 (1955).
57. ——, "Some Aspects of Industrial Diversification in the United States," *Economic Geography*, Vol. 33 (Jan. 1957).
58. Schooler, Eugene W., *Regional Advantage in the Production of Chemicals from Petroleum and Natural Gas,* doctoral dissertation, Harvard University, Cambridge, Massachusetts, 1954.
59. Schurr, Sam H., and Jacob Marschak, *Economic Aspects of Atomic Power*, Princeton University Press, Princeton, New Jersey, 1950.
60. Shapiro, G., "Myrdal's Definition of the South: A Methodological Note," *American Sociological Review*, Vol. 13 (Oct. 1948).
61. Shevky, Eshref, and Wendell Bell, *Social Area Analysis*, Stanford University Press, Stanford, California, 1955.
62. Smith, Thomas R., *The Cotton Textile Industry of Fall River, Massachusetts*, King's Crown Press, New York, 1944.
63. Stouffer, Samuel A. et al., *Measurement and Prediction*, Princeton University Press, Princeton, New Jersey, 1950.
64. Thompson, Wilbur R., "The Coefficient of Localization: An Appraisal," *Southern Economic Journal*, Vol. 23 (Jan. 1957).
65. Thünen, Johann Heinrich von, *Der isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie*, Hamburg, 1826.
66. Thurstone, L. L., *Multiple Factor Analysis: A Development and Expansion of the Vectors of Mind*, University of Chicago Press, Chicago, 1947.
67. Van Arsdol, M. D., Jr., S. F. Camilleri, and C. F. Schmid, "An Application of the Shevky Social Area Indexes to a Model of Urban Society," *Social Forces*, Vol. 37 (Oct. 1958).
68. Weber, Alfred, *Theory of the Location of Industry*, translated by C. Friedrich, University of Chicago Press, Chicago, 1929.
69. Zelinsky, W., "A Method for Measuring Change in the Distribution of Manufacturing Industry: The United States, 1939–47," *Economic Geography*, Vol. 34 (April 1958).

Chapter 8

Interregional and Regional Input-Output Techniques*

A. Introduction

It is commonplace among those who wish to embark upon regional studies to ask what kind of study would be most fruitful, given their set of objectives and terms of reference. As they scan the list of possible types of studies, such as regional income studies, commodity flow studies, balance of payments studies, economic base studies, multiplier studies, industrial location studies, etc., they are perplexed as to which one to attempt. They see the virtues and limitations of each. Often they clearly perceive the partial character of each and are dissatisfied. They want to study more of the whole of the region or of the system of regions. Or if they have the resources to do more than one partial study, they seek ways by which these studies may be interrelated and may be conceived and formulated to contribute to one another.

It is in this connection that the general interdependence techniques are of value. These techniques have many limitations: they involve sweeping assumptions; they abstract from many important realities of folk, regional, national, and international life. Yet, after all their limitations are set down, the fact remains that they provide essential mortar for cementing various partial studies.

* This chapter was written with John H. Cumberland.