OPTIMUM CITY SIZE: FACT OR FANCY?

MALCOLM GETZ*

Before a policymaker can sensibly consider whether or how to control urban growth, he must first ask whether there is one city size that is optimal for achieving the highest level of overall welfare. If one city size is preferable to another, the next question is whether decision-making through decentralized markets will lead to the desirable city size, or, conversely, whether some affirmative interventionist strategy is necessary to achieve a more desirable size. Finally, if decentralized markets are likely to be unreliable in obtaining a socially desirable outcome with respect to city size, the problem is to identify strategies of government intervention that can effectively influence city size. This article will consider each of these questions in turn.[1]

* Associate Professor of Economics, Vanderbilt University.

1. The notion of a city here is an economic not a political concept; a city is considered to be a densely developed area. For a variety of reasons, political boundaries of a city seldom conform to the economic definition. The Bureau of the Census makes a detailed measurement of urbanized areas, defining a Standard Metropolitan Statistical Area (SMSA) as all those contiguous counties encompassing the urbanized area. Because much information is available about SMSAs, the SMSA is the most useful approximation to the economic concept of a city. An area thus defined must contain a resident population of 50,000 in order to qualify as an SMSA. There were about 276 SMSAs in 1976, and all states except Wyoming and Vermont contain at least some part of an SMSA. SMSA's are listed in STATISTICAL ABSTRACT OF THE UNITED STATES (1977), at 923-31.

For purposes of discussing optimal city size, however, the definition at issue is serious and consequential. Problems occur for both the smallest and the largest places. Could one study the distribution of city sizes by looking only at SMSAs while ignoring smaller places? In 1790, only three cities might have satisfied current definitions of an SMSA and so a discussion of their sizes might ignore a lot of the relative differences in the sizes of developed places. In 1970, on the other hand, 68.6 percent of the population of the United States lived in SMSAs, so looking at SMSAs now seems to capture most of the issue. STATISTICAL ABSTRACT OF THE UNITED STATES (1975), at 17, Table 17.

Only 82 percent of the urban population of the country lives in SMSAs; the other 18 percent live in urban places too small to be considered SMSAs, *id.* at 17, Table 17. The notion of an optimal city size may have applicability to such smaller cities, if it could be determined that there exists some critical minimum city size where the quality of life is significantly better than in smaller places.

Definitional problems also arise for large cities. If an urbanized area grows so that previously separated cities become joined, the contiguously urbanized area may be very large indeed. The Census Bureau has recognized such megalopolises by defining Standard Consolidated Areas (SCAs). Currently, the New York SCA includes New York City, Nassau-Suffolk, and Northern New Jersey SMSAs, and the Chicago SCA includes Chicago and Gary-Hammond-East Chicago. One might expect an SCA to be defined in Southern California soon. Should cities that have grown together be considered as single cities or should they be treated as separate? As will be seen in the following discussion, we are interested, for example, in the impact of city size on air quality, crime, commuter travel time, and consumption opportunities. The relevant concept of city should be the one where the connection between size and the quality of life is most meaningful. Of course the distinctions are not very tidy, perhaps not tidy enough for contemplating public policy. If we are to consider defining an optimal city size, we must acknowledge and confront the definitional issue.

I
AN OPTIMUM SIZE DISTRIBUTION OF CITIES

A. Why Cities at All?

The history of cities gives strong evidence that their growth stems from the development of technologies that require large-scale production and from the growth of markets large enough to take advantage of these new technologies. If a single individual could produce goods as efficiently as workers who specialize in the production of a single good, then there would be no cities. City life involves transportation and organizational costs, costs of shipping food and fiber to the city, costs of removing waste from the city, and costs of agreeing on, imposing, and enforcing complex rules for living. Thus, if there are few advantages to production in cities sufficient to justify the extra expense, few will choose city life.

But economies of scale in production appear very early[2] and the development of cities matches changes in technology that create advantages for large-scale production.[3] As technological change and economic opportunity created advantages for large-scale production of goods and services, cities developed to take advantage of these opportunities.

The advantages of size, which led to the development of cities in the first place, must be balanced against the added costs of maintaining city life. Technology influences the determination of a desired city size not only with respect to production economies but also with regard to coping with the costly disadvantage of city living. For example, improvements in transportation technology have made the movement of goods and people easier within the city and between the city and its hinterland. Similarly, modern water and sewage systems reduce the disadvantages of urban life. These technological developments made city life more manageable, enhancing urban growth and allowing the desirable city size to grow larger.

Consideration of the advantages and disadvantages of city size and possible means for coping with the disadvantages allows articulation of a simple notion of optimum city size—namely, that where the marginal social benefit from increased production scale is just equal to the marginal social cost of the additional increase in city size required. This rule defines an optimum when the additional benefits diminish and the extra costs increase for a larger city.[4]

---

2. For example, irrigation and defense systems caused cities to appear very early along the Tigris and Euphrates Rivers as production activities involving advantages to large scale production arose. H. HEATON, ECONOMIC HISTORY OF EUROPE (1948), at 13-14.

3. Production must include the production of knowledge, culture and government involving changes in scale that may require the existence of cities if they are to be produced.

4. J.A. Mirrlees, *The Optimum Town*, 74 SWEDISH J. OF ECON. 114 (1972); A. Dixit, *The Optimal Factory Town*, 4 BELL J. ECON. & MANAGEMENT SCI. 637 (1973).

B. Hierarchy and Hinterland

The economy is composed of myriads of production activities, each potentially having different scale economies. Therefore, an optimum city size for one production activity may be inappropriate for others. Instead of a single optimum city size, there may be an optimum distribution of city sizes, with each city sized according to the scale required by the production processes within it.

If each city serves its hinterland, a relationship will exist between the size of the hinterland and the scale of activities within the city. For small hamlets, the economic activities will be of small scale: convenience food shops and vehicle service stations. The hinterland or market area for these activities is small. Progressively larger towns provide services on a progressively larger scale and require a larger market area: vehicle dealers, furniture stores and specialty shops. Only the largest cities with the largest market areas can provide services that involve the largest scale economies: investment banks, business services, and wholesaling functions. Thus, the optimum city size distribution may be an urban hierarchy.[5]

---

5. The notion of cities as central places serving market areas was developed in W. CHRISTALLER, CENTRAL PLACES IN SOUTHERN GERMANY (C. Baskin trans. 1966). The hierarchy theory was developed in A. LOSCH, ECONOMICS OF LOCATION (F. Wolfgang trans. 1954).

George Zipf has observed an empirical regularity in the size distribution of cities—namely, that the rank of the city (beginning with one for the largest city) multiplied by the population of the city tended to be a constant for all cities in a country. ZIPF, HUMAN BEHAVIOR AND THE PRINCIPLE OF LEAST EFFORT (1949). The rank-size rule, or Zipf's Law, was linked to the more formal theory of the urban hierarchy in Beckman, *City of Hierarchies and the Distribution of City Size*, 6 ECON. DEV. & CULTURAL CHANGE 573 (1961). If a city of a certain size is to serve a given rural area, and a larger city can serve a limited number of cities of the next smaller size, then city size increases geometrically with the rank of the city from smallest to largest.

While the rank-size rule finds some empirical re-enforcement, it is ultimately of little policy value. The rank-size rule has been found to fit the distribution of cities over size in the United States and seems to fit many other situations at least roughly. E. MILLS, URBAN ECONOMICS 106-08 (1972); B. BERRY, GEOGRAPHY OF MARKET CENTER AND RETAIL DISTRIBUTION (1967). The rank-size rule fits least well in countries where one city dominates, a phenomenon called primacy; Buenos Aires, Karachi, Nairobi and Copenhagen are examples. Berry has explored the relationship between the size distribution of cities and the level of economic development and finds little association. Berry, *City Size Distributions and Economic Development*, 9 ECON. DEV. & CULTURAL CHANGE 573 (1961); Berry, *An Inductive Approach to the Regionalization of Economic Development*, in ESSAYS ON GEOGRAPHY AND ECONOMIC DEVELOPMENT 78 (N. Ginsburg ed. 1960). Small countries, countries where foreign trade is important, and countries where government, industry and trade are centered in a single city are more likely to be primate.

The existence of primacy indicates the weakness of the theory underlying the rank-size rule. In particular, the theory is too casual in defining the area where the rank-size rule should apply to the distribution of cities. In some sense an economic trading area is a more appropriate area than national boundaries, yet no specific method is proposed for defining a trading area separate from political boundaries. Thus the rank-size rule does not give clear indication of the optimality of a particular distribution of city sizes.

The rank-size rule and the urban hierarchy theory seem inappropriate in an industrialized economy. The hierarchy hitches the size of cities to the size of smaller cities and hinterlands. The implied trade flows are between hinterland and city, between small and large cities. Yet, in an in-

That cities face increasing marginal costs as they become larger and that there are substantial economies of scale in production in many industries suggest that cities could specialize in production of a few goods, so that a city could be as small as possible while taking advantage of scale economies. The advantages and disadvantages of city size may occur for both production and consumption. Each of these is discussed below.

C. Production and City Size

The productivity of economic activity may differ across cities because of differences in plant size, industry size, in city attributes, and innovation. If an industry is represented in a city by a single plant, then the local scale of the plant and of the industry are the same. Economies of scale in an individual plant or firm may be simply technological in nature: a minimum size assembly line may be feasible. Steel mills, vehicle assembly and airplane manufacturing plants are examples of large-plant industries that may dominate a single city.[6] Governmental centers dominate Washington, D.C. and Sacramento.[7] Most cities, however, are not dominated by a single firm.[8] Thus, economies of scale within a plant or a firm are not sufficient to explain city size.

Industry size within a city may generate production advantages. Individual firms may specialize. A specialized labor pool might attract a particular industry by enhancing recruitment opportunities, or the availability of specialized business services may make such in-house services unnecessary. Market and technical information may be more readily available in an industrial city allowing production plans to be revised and adopted more quickly. Economies of scale of an industry that are external to the firm are called agglomeration economies. The classic example, discussed by Benjamin Chinitz, is the fashion industry in New York.[9] Small firms make specialized products: sleeves, buttons, fabrics, etc. Quick response to a rapidly changing market is critical to success. Financial institutions are attuned to working with volatile fashion-

---

dustrialized economy, the cities tend to specialize in the production of particular goods and services for export to other cities. Thus, the most important flows may be among cities of roughly comparable size. Trade between New York and Chicago is as important as trade between New York and Trenton. Cities of similar size are not producing a similar market basket of goods for export to smaller or larger cities as suggested by the hierarchy theory. Knowing only a city's size, one knows very little about what might be produced for export in that city. Therefore, the urban hierarchy theory is irrelevant to understanding city size in an industrial economy and of no value in identifying an optimum distribution of city sizes.

6. *E.g.*, Bethlehem Steel in Bethlehem, Pennsylvania; U.S. Steel in Birmingham, Alabama and Duluth, Minnesota; and American Motors in Kenosha, Wisconsin.

7. Of a work force of nearly 1.3 million in Washington, D.C., 39.1 percent were employed by government in 1972. For the same period, the figure for Sacramento, California was 34.7 percent. U.S. BUREAU OF THE CENSUS, COUNTY/CITY DATA BOOK 1972, 560, 570 (1973) [hereinafter cited as DATA BOOK].

8. *Id.* at 550.

9. *See* Chinitz, *Contrasts in Agglomeration: New York and Pittsburgh*, 51 AM. ECON. REV. (May 1961) at 279.

apparel firms. Entry of new firms is easy since no single firm need play more than a minor role in turning cloth into profits. The agglomeration economies within an industry seem to represent the economies of information flows where frequent, face-to-face contact is important to manufacturing, marketing and finance.

Examples of cities where a single industry dominates are somewhat better than those for a single plant, but the evidence is far from compelling. Pittsburgh and Birmingham may be dominated by steel, Detroit by automobiles, Rockford by hardware, Hartford by insurance, but these cities have diversified exports so that agglomeration economies within the dominant industry alone cannot explain city size. Larger cities—New York, Boston, Chicago, Philadelphia, and Los Angeles—are extremely diversified and were not shaped solely by agglomeration economies within a single industry.[10]

If firm and industry scale economies are inadequate to explain production concentration in large cities, then one must look to interindustry serendipity. Large industry suppliers and buyers may find advantages (e.g., in communication or transportation) in locating near their buyer or supplier. Interindustry purchases and sales may encourage small firms to locate in large cities. Steel fabricators could choose Pittsburgh and automobile parts manufacturers, Detroit.

Interindustry effects are not limited to manufacturing. Access to wholesaling, parts, or financial markets may be significant. These non-industry-specific agglomeration economies are difficult to substantiate in a detailed manner. Two essays examine the productivity of cities: Leo Sveikauskas correlates value-added per worker to city size, the population's education, and regional dummy variables, and found significant positive associations between productivity and city size in eleven of the fourteen two-digit manufacturing industries he examined.[11] The difference in productivity is greater than can be explained by differences in capital per worker. Sveikauskas concludes that many industries are more productive in larger cities.[12] David Segal correlates output to labor and capital in cities, controlling for human capital differences.[13] Segal's data on capital for each city do not allow separate consideration of industries, but do allow for estimation of the degree of economy of

10. *See* DATA BOOK, *supra* note 7, at 550, 560, 570.

11. Sveikauskas, *The Productivity of Cities*, 89 Q. J. ECON. 393 (1975). Two-digit industries are defined in: OFFICE OF MANAGEMENT AND BUDGET, STANDARD INDUSTRIAL CLASSIFICATION MANUAL 1972 [hereinafter cited as SIC] (1973), at 9-14. The first digit of the SIC code designates major categories of economic activity: agriculture, mining, construction and the like. The second digit describes a somewhat more detailed classification. In manufacturing (*see* SIC, at 57-218), for example, the following are examples of two-digit industries: SIC 21 tobacco, SIC 22 textile mill products, SIC 25 furniture, SIC 26 paper and allied products. When a third- and fourth-digit are used, still more detailed categories of activity are defined, *i.e.*, SIC 2642 envelopes and SIC 2643 paper bags, *see* SIC 100-02.

12. Sveikauskas, *supra* note 11, at 395-96.

13. Segal, *Are There Returns to Scale in City Size?*, 58 REV. ECON. & STATISTICS 399 (1978).

scale. He finds no evidence of scale economies of production but does find that cities over two million in population are about eight percent more productive than cities under two million.[14] Segal concludes that there is evidence of agglomeration economies in cities over two million. Neither study details the nature of the agglomeration economies. This is an important area for future research on optimal city size.

Other city attributes than size may influence productivity. Firms may value public services, e.g., sewer and water pricing, power availability, local tax breaks, public assistance in labor training, or special site development concessions. They may also consider public policies on crime and air quality, although the relationship between city size and these factors is not well defined. Similarly, climate and labor force characteristics may be important to some firms. The question of firm productivity in a city seems more complicated than Sveikauskas and Segal allow. Productivity studies must consider dynamic aspects: What is the technological rate of change for small, medium and large cities? Can we make generalizations about the filtering of ideas from one city size to another? These questions await further study. Overall, there is little convincing evidence that some city sizes are better than others. Firm-industry-interindustry scale effects fail to explain the differences in city size. General city agglomeration economies in production are difficult to substantiate in the detailed manner that is necessary. The impact of city size on innovation seems impossible to gauge.

D. Consumption and City Size

Human welfare is determined by more than the ability to earn and spend income. It is determined increasingly by environmental factors that cannot be purchased directly in a market[15] and vary from city to city. City size affects environmental quality. It is possible that future cities' optimal size will be determined by environmental factors.

Quality of life in cities is multi-dimensional. The dimensions have to be added and each weighted according to its impact on human welfare. Early studies of the quality of life used primitive methods to add dimensions and did not weight them.[16] Recent researchers have estimated hedonic prices[17] for

14. *Id.* at 339.

15. J. Tobin & W. Nordhaus, Economic Growth (1972) (Retrospect and Prospect Fiftieth Anniversary Colloquium V, National Bureau of Economic Research).

16. B. Liu, Quality of Life Indicators in the U.S. Metropolitan Areas, 1970 (1975) (Midwest Research Institute report, Kansas City).

17. Consumer goods usually have a variety of attributes. A refrigerator for example has size, freezing capacity, color and a rate of electricity consumption for a given kind of use. In buying a refrigerator, one pays one price but receives a collection of attributes. The extra cost of buying a more efficient refrigerator with other attributes the same is an implicit price for some unit of energy efficiency in the refrigerator.

As consumers choose among all the refrigerators available in the marketplace, the implict prices of the attributes of the refrigerator in equilibrium will reflect the individual attributes of

the individual attributes and, using the prices, constructed an index of the quality of life.[18] In a hypothetical world where families are mobile between cities and mobility costs are low, allowing families to relocate quickly in response to new opportunities,[19] we assume that the quality of life in each city is relatively homogeneous (for individuals of particular incomes and occupational groups). In this world, consumers must like the city where they live, otherwise they would move. Given perfect mobility, any differences in the quality of life in various cities would be reflected in real earnings differences (monetary earnings divided by the cost of living). The marginal value consumers assign to quality of life factors might be estimated by examining how real earnings vary for different qualities of life. Given sufficient variation in attributes across cities, hedonic prices may be estimated for each attribute. Such an exercise must allow for possible differences in the human capital of the work force, in productivity of workers, and for the possibility that families may not adjust rapidly to changing opportunities. It must also include all the relevant attributes of the quality of life.

Getz and Huang have studied the relationship between earnings and the quality of life across cities.[20] They examine earnings in nine urban occupational groups so that differences in occupational mix from city to city do not have an undue effect on earnings.[21] The age and education of the adult pop-

---

both the extra production costs of changing the attribute (supply) and the marginal consumer's valuation of the attribute (demand). The marginal consumer valuation of the attribute is the hedonic price of the attribute. *See* Rosen, *Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition*, 82 J. POLITICAL ECON. 34 (1974).

Urban amenities will be some of the attributes of location important to consumers as they choose cities and residences within cities. In equilibrium, market prices will reflect the hedonic prices of the urban amenities. Within a city, variation in property values may reflect differences in urban amenities. *See, e.g.*, Nelson, *Residential Choice, Hedonic Prices, and the Demand for Urban Air Quality*, 5 J. URB. ECON. 357 (1968). As consumers choose among cities, the wage rate in different cities may reflect the hedonic prices of more aggregate urban amenities.

18. O. Izraeli, Differentials in Nominal Wages and Prices Between Cities (1973) (Urban Economics Report #71, University of Chicago). Hoche & Drake, *Wages, Climate, and the Quality of Life*, 1 J. ENVT'L ECON. AND MANAGEMENT 268 (1974); S. Rosen, *Wage-Based Indexes of Urban Quality of Life*, in CURRENT ISSUES IN URBAN ECONOMICS (P. Mieszkowski & M. Straszheim ed. 1979), at 74-104; Getz & Huang, *Consumer Revealed Preference for Environmental Goods*, 60 REV. ECON. & STATISTICS 449 (1978) [hereinafter cited as Getz & Huang].

19. Differences in earnings in different cities will reflect the hedonic prices of urban amenities only if the cost of moving among cities is small relative to the lifetime earnings stream. 36.9 percent of the population moved between 1970 and 1974. Of these, 5.5 percent moved between Standard Metropolitan Statistical Areas, 2.1 percent entered an SMSA, 3.1 percent left an SMSA without entering another and 8.8 percent moved with neither origin nor destination in an SMSA. U.S. Bureau of the Census, *Current Population Reports: Population Characteristics*, Dec. 1974, at 12. Relatively small rates of migration among cities may be sufficient to make utility levels near equal in different cities. Persons entering the labor force or who are working may be among those most likely to move in order to better their situation, although all groups of the population seem to be quite mobile.

20. *See* Getz & Huang, *supra* note 18.

21. *Id.* at 451. The implicit valuations of urban amenities may differ for different groups. Operatives may be more concerned with employment opportunities for females while professionals are more concerned with consumption amenities or the quality of schools. By estimating separate

ulation are considered in order to control for differences in human capital. Factors shaping production are considered so that differences in productivity do not affect the estimated relationship.[22] Net migration is considered in order to account for the adjustment process.[23] Finally, a wide variety of life quality measures is included.[24]

Getz and Huang find that earnings vary across cities in keeping with differences in the quality of life. Effects such a crime, environmental quality, variance in commuter travel time, and variety of consumer opportunities vary with city size and are difficult to isolate, but as a group, they are important.[25] Variation in earnings associated with climate and region are slight. One consistent finding is that earnings of males vary inversely with the employment opportunities of females. That is, males accept lower earnings where females are more likely to find employment. This effect is larger in lower than in higher income occupations.[26] Huang weights consumer amenities that vary with city size with consumer values and sums them, assigning a net value to amenities in particular city sizes. From his evidence, Huang concludes that the highest net value of amenities is found in cities with a 250,000-500,000 popu-

---

hedonic prices for different occupational groups, Getz and Huang discover some differences in tastes for amenities by different occupational groups.

22. *Id.*

23. *Id.*

24. *Id.* at 451-52. The characteristics of life in a city that may be viewed as amenities by households in choosing a location are quite varied. Climate, the chance of being a victim of crime and the quality of air are frequently considered. The average distance a commuter must travel to work varies across cities. While variation in the expenditures for commuting are reflected in the cost of living index, the greater inconvenience and value of time spent commuting is not. Large cities have more restaurants, more cultural opportunities, more kinds of retail shops. More, however, is not more of the same but a greater variety. The value of the extra variety is not reflected in prices. Therefore, the greater variety of consumption opportunities may be valued as an amenity by consumers. The taxes paid locally are reflected in the cost of living index, but the quality of services for given tax dollars may differ. The net value of services of tax payments is called a fiscal residual. The value of fiscal residuals in education, health services, refuse collection and the like may be viewed as amenities that vary over cities. Finally, the character of job opportunities may be viewed as an important amenity. Households may seek a joint utility maximization and so the relative availability of jobs held by men and women may be seen as one determinant of the quality of life.

Getz & Huang, at 453-457, regress earnings in each occupational class on the cost of living, the net migration rate and a broad set of environmental attributes. A two-stage estimation technique is used to control for differences in productivity in a city (demand characteristics) so that the coefficients of the relationships can be interpreted as the consumer's marginal valuation of the attributes.

25. *Id.* at 454.

26. *Id.* at 454-57. One part of the consumer environment that bears special mention is the public sector. Swanson, Smith and Williamson develop a notion of an optimal city size from the perspective of providing public services. Swanson, Smith & Williamson, *The Size Distribution of Cities and Optimal City Size*, 1 J. URB. ECON. 395 (1974). In the Getz & Huang study taxes are reflected in the cost of living and the value of public services is, in principle, included as a non-purchased amenity. The quality of health services and the quality of education are central examples. A dollar of taxes may generate services of different values in different places. This fiscal residual may be a consequence of economies or diseconomies of scale in public services. The fiscal residual from different services is just one class of non-purchased amenity of city living.

lation range. The net value lowers for larger or smaller sized cities and reaches a minimum at five million, increasing thereafter. These studies suggest that it is possible to estimate precisely the valuation consumers place on attributes of the quality of life. More precise estimates could be used to determine how the quality of life is valued in dollars for different cities.

If productivity is higher in cities that consumers find unattractive, how can we determine the optimum city size? Trade-offs between conflicting objectives is the essence of economics. The stock economic answer to this question is that prices could be adjusted, allowing consumers to receive higher wages for living in less attractive areas. Producers would pay the higher wages for the higher productivity payoff. The cost of land may also vary so that firms will choose the high rent areas only when estimated high productivity justifies the choice. Only in terms of relative productivity attractiveness can we understand optimum city size principles. The public policy question is whether decentralized location decisions of households and firms lead to cities of optimum sizes.

II
CITY SIZE POLICY

For an economy to allocate its resources efficiently by relying solely on decentralized markets, each economic unit must bear the full consequences of its decisions. If some consequences spillover to others without their consent (i.e., if externalities are present), it is likely that resources are being misallocated. Since spillovers are more common when economic activity is more dense, unregulated decentralized markets are unlikely to allocate resources efficiently in urban areas.[27]

When market failure occurs, the first impulse for improving resource allocation efficiency might be to try to simulate an economy where the requirements for effective decentralized markets exist. Market prices might be adjusted so that each economic agent bears the full consequences of its action. For example, taxes or charges might be imposed so that a congested street

27. Yuh-ching Huang, Consumers' Revealed Preference and Social Optimum Distribution of Urban Sizes 37 (1975) (Ph.D. dissertation, Vanderbilt University).

For example, a city street user incurs vehicle costs. When the street is crowded, the user imposes extra costs on other street users due to slower traffic flow. Other drivers do not bargain for additional congestion, thus some street-use costs affect others without their consent. Therefore, the street will be over-utilized and too congested because no user is required to bear the full cost of his/her activity. Similarly, pollutant emissions affect others in ways for which they have not bargained. Decentralized markets—as in cities—will not yield an efficient allocation of resources where such spill-overs are important.

Spill-overs need not be negative. If economies of scale exist, an additional consumer may have the effect of lowering costs to all consumers. This benefit will not accrue uniquely to that incremental consumer. Such scale effects are not limited to plants or firms, but may extend to whole industries or cities. A larger buyer pool creates greater consumer variety. The benefit of additional variety will accrue to all consumers together. Because the individual consumer does not bear the full consequences of the change in consumption opportunities following his/her relocation, resources are likely to be inefficiently allocated.

user must bear not only his individual costs, but also a charge reflecting the extra travel time costs imposed on other street users. Charges could be imposed on air pollution emissions, putting responsibility on the polluter to bear the costs imposed on others by the emission. If these external diseconomies were the only causes of market failure and all externalities could be charged in a simple, low-cost method, then human welfare could be greater.[28]

The relationship between city size resulting from full social-cost pricing—charging for pollution, congestion, and the like—and market equilibrium in the absence of such pricing is ambiguous. George Tolley asserts that pricing out externalities would encourage more efficient use of space and, in some circumstances, the efficient city would be larger than the market result.[29] James Henderson finds that in a variety of cases, the optimal city employing congestion charges is larger than a city where streets are financed

28. Garrett Hardin, in his trenchant article on the tragedy of the commons, illustrates the problem of social control that stems from a similar circumstance. He describes a commons used by shepherds for feeding their flocks. Each shepherd secures the full benefit of grass used to feed one of his flock. At the same time, each time grass is used up, every shepherd receives a disbenefit because of the reduced supply of grazing land available for feeding. However, whereas a shepherd secures the total benefit from using land to feed his herd, he shares the disbenefit evenly with the other shepherds, and therefore any given shepherd has an incentive to add to the herd provided there remains a part of the commons still available. Clearly, however, this system has within it the seeds of its own destruction since no single decision-maker has any incentive to consider the disbenefits to the community that flow from increased utilization of the commons. As Hardin poignantly points out, decentralized decisionmaking in a commons can result in disastrous, self-destructive consequences. *See* Hardin, *The Tragedy of the Commons*, 162 SCIENCE 1243 (1968).

29. The pollution example is a classic. If a polluting firm considers costs of production in determining prices and output, it is unlikely to include a calculation to cover the costs imposed on pollutees—those who suffer at the expense of the firm's pollution without having given their consent. As a consequence, the perceived costs of the firm—the so-called private costs—will be lower than the "real" social costs, which must include the (external) effects on the pollutees. This understatement of costs is likely to result in an excessive allocation of resources to the goods produced by the polluting firm, whose cost function ignores an important component of overall social cost.

The pollution example also demonstrates the importance, from a distributive perspective, of the allocation of legal entitlements. Presumably, where transaction costs and administrative costs are low, society can require a polluting firm to compensate unconsenting pollutees according to an objectively determined measure of damages. *See, e.g.*, Boomer v. Atlantic Cement Co., 26 N.Y.2d 219, 257 N.E.2d 870 (1970). This allows a firm to pollute but requires that, by paying damages, it internalize the costs it imposes on unconsenting victims. This form of legal rule assigns a legal entitlement of the pollutees, protected by a liability rule—*i.e.*, damages. A different form of legal protection would be afforded by a property rule, where the pollutee's legal entitlement would be protected absolutely against impairment and a prospective pollutee could sue under a nuisance theory to enjoin the polluter. *See* Calabresi & Melamed, *Property Rules, Liability Rules, and Inalienability: One View of the Cathedral*, 85 HARV. L. REV. 1089, 1092, 1115-21 (1972).

It is clear that the assignment of entitlements and the form of legal protection accorded those entitlements will influence the distribution of wealth and, where transaction costs exist, may affect the allocation of resources. *See, e.g.*, Coase, *The Problem of Social Cost*, 3 J. LAW & ECON. 1 (1960); Demsetz, *Wealth Distribution and the Ownership of Rights*, 1 J. LEGAL STUDIES 223 (1972).

with gasoline excises.[30] Edwin Mills and David deFarranti, in their discussion of pollution and congestion as sources of market failure, suggest that the optimal city size may be larger or smaller than the optimum if all externalities were appropriately priced.[31] It is generally agreed that market equilibrium city size may differ from the optimum, but it also seems that the optimum may be smaller or larger than the equilibrium depending upon the specific characterisitics of the market failures. If all externalities could be priced appropriately to yield the efficient allocation of resources, then decentralized markets could be relied on to achieve a socially desirable outcome for city size. Effectuation of such a system seems unrealistic. Such charge systems are complex and would be expensive to administer, and placing a value on externalities is difficult. Historically, public policy has coped with externalities by compromising between efficiency and effectiveness. Treatment requirements or limits on individual automobile emissions are likely to be inefficient methods of improving the environment.[32] They are enacted because they are simple to understand and enforce. City size policy may have the same appeal. City size has an influence on traffic congestion, environmental quality, crime, productivity, and consumer opportunities. Perhaps human welfare could be improved by gentle manipulation of city size when detailed efforts to deal with the externality issues are inadequate. If it is impossible to charge externalities to consumers so that consumers bear full responsibility for their ac-

30. Various approaches to environmental policy are discussed in Oates & Baumol, *The Instruments for Environmental Policy*, in ECONOMIC ANALYSIS OF ENVIRONMENTAL PROBLEMS 95-128 (E. Mills ed. 1975) [hereinafter cited as Oates & Baumol].

31. A series of theoretical works deal with the relation between an optimum city size that maximizes the utility of the residents and an equilibrium size that is the result of market forces. If there were no externalities or economies of scale, the optimum and equilibrium sizes would be the same. George S. Tolley considers negative externalities like air pollution. In the equilibrium situation, the externalities are unpriced and so over-produced. In the optimum situation, the externalities are internalized through appropriate prices. Tolley finds that if the negative externalities occur in the production of goods for export from the city in national markets with given prices, internalizing the externality will make the city smaller; that is, the optimum city size is smaller than the equilibrium. On the other hand, if the externality is produced by firms producing local or non-traded goods, the optimum will be larger than the equilibrium. Tolley, *The Welfare Economics of City Bigness*, 1 J. URB. ECON. 324 (1974). James V. Henderson discovers that moving from gasoline tax financing of roads to optimum congestion charges plus a head tax tends to increase travelling speeds and lower commuting costs for a plausible range of values. Thus, internalizing the congestion externality would seem to make the optimum larger. Henderson, *Road Congestion: A Reconsideration of Pricing Theory*, 1 J. URB. ECON. 346 (1974). Henderson considers optimum and equilibrium city size explicitly. In this model the amount of space devoted to roads is taken as endogenous, that is, chosen efficiently given land rents and travel costs. Some economies of scale are allowed in the provision of the road. The Pareto-optimum city size will again be larger than the equilibrium size; when optimum congestion charge rents are lower, more space is devoted to roads and higher travelling speeds are achieved with lower total commuting costs. Note that the Henderson findings are consistent with Tolley's: congestion affects a non-traded good and so internalizing it tends to make the city larger. Henderson, *Congestion and Optimum City Size*, 2 J. URB. ECON. 48 (1975) [hereinafter cited as Henderson, *Congestion and Optimum City Size*].

32. Henderson, *Congestion and Optimum City Size*, *supra* note 31.

tions, then it may be appropriate to push city size toward some target size (or set of sizes) that represents an improvement in human welfare over the current distribution of city sizes. That outcome would not necessarily constitute an optimum which would arise if externalities were fairly and accurately priced. Rather, it would reflect an estimate based on empirical data and might well be inefficient, even though an improvement over the existing situation.

A straightforward policy to shape the city size distribution could be justified on the grounds that it is administratively a simple method to encourage the production of valuable externalities and discourage negative externalities by inducing some firms and households to elect to locate in cities where the net value of externalities is greatest. If one could estimate accurately the relationship between city size and human welfare, then a tax and subsidy scheme could be devised along the following lines: Suppose a metropolitan population increases over one million, thus increasing consumer opportunities, amenities, and worker productivity, but increasing disamenities to a greater extent. Because the marginal contribution of each new resident is greater than the average (the average is what each consumer considers in response to a location decision), then an earnings tax reflecting the difference between marginal and average contribution to environmental change would force the wage earner to confront an approximation to the marginal social cost of selecting a particular city. Similarly, if cities under 250,000 create opportunity amenities or worker productivity increases much more rapidly than commuter travel time, deterioration in air quality, and the like, then a subsidy to wage earning in such cities might be appropriate to encourage their growth. Such a tax and subsidy scheme might be keyed to the Social Security payroll tax. That is, the payroll tax might be linked to and vary with city size.[33]

A payroll tax linked to city size would encourage workers and firms to choose to locate in cities where additional workers might raise the quality of life by enhancing the variety of consumer opportunities and increasing the benefits of scale and agglomeration economies in production. Similarly, the tax would discourage workers from choosing to move to cities where the marginal contribution of an additional worker to the quality of life would be negative. There is no claim that such a variable rate payroll tax would be very precisely associated with the externalities leading to market inefficient city sizes. Indeed, the resulting distribution of city sizes may differ from the no-externality result. Rather, such a tax is simple to administer and may improve human welfare when direct approaches to externalities fail. The confidence one might have in such a proposal is directly related to the confidence one has in the estimated relationships between city size and the quality of life. The

33. Mills & deFerranti, *Market Choices and Optimum City Size*, 61 AM. ECON. REV. 340 (1971).

estimates to date are too speculative to warrant experimentation with such a tax.

If the direct approach to shaping the size distribution of cities seems unwarranted given current information, does the notion of an optimal city size have any policy value now? Present federal taxes and expenditures probably influence the size distribution of cities. General revenue sharing and a variety of special revenue sharing programs disburse funds differentially based on urbanization. Federal transportation aid[34] (especially funds for mass transit) are concentrated in the largest cities.[35] Comprehensive Employment and Training Act[36] funds are concentrated in large cities. Of course, many

---

34. J. KRIER, ENVIRONMENTAL LAW AND POLICY 300 (1971). Wallace Oates and William Baumol compare tax, subsidy, suasion and direct controls as methods of dealing with pollution. They suggest that direct controls may be desirable where the results of other policies may be unpredictable. *See* Oates & Baumol, *supra* note 30. They agree, however, that the pricing system has been too little used in the design of environmental policy. In particular, it is important to compare the incentives for technological change in environmental control techniques when comparing direct controls with priced-based schemes. Allen V. Kneese and Charles L. Schultze discuss current environmental legislation and indicate how firms will be penalized for developing new techniques to control pollution under the current direct control policies. A. KNEESE & C. SCHULTZE, POLLUTION, PRICES AND PUBLIC POLICY (1975). The direct control approach also involves substantial costs in terms of the forced timing of changes. Pricing schemes, on the other hand, create a continual incentive for technological change allowing firms to make their own decision about the timing of changes in technique. Over time, the direct control approach may be very costly indeed.

35. The tax system may not currently be strictly geographically neutral. The progressive bracket structure of the personal income tax collects more taxes from cities, where the cost of living is higher, because earnings are higher there to compensate for the cost of living. The Social Security payroll tax might be modified to include geographic differentials in rates without changes in benefits. Perhaps differentials would be limited to the employers' contribution. If areas where growth would generate more external economies than diseconomies could be identified, then lower payroll tax rates there might induce further beneficial development. In other areas, growth might increase external diseconomies more than external economies and so a higher payroll tax might be appropriate to discourage development in such places. Payroll tax rates might vary by county, for example, to reflect the marginal impact of growth on externalities.

The geographic differentials in taxes cause changes in land rents in different counties. Thus, the distributional impact of the tax would be on land owners, with land values in underdeveloped areas increasing and land values in overdeveloped areas decreasing. Such changes in land values would likely take some time to develop as firms and households take time to consider relocation. Thus, the distributive impact of the geographic tax differential will be superior to that of many other growth modifying proposals because it applies to all land and workers, not just to those who happen to want to move or to develop land.

36. The Advisory Commission on Intergovernmental Relations reports a weak association between the distribution of categorical grants and income, rural areas and smaller populations. Advisory Commission on Intergovernmental Relations, *Categorical Grants: Their Role and Design*, in THE INTERGOVERNMENTAL GRANT SYSTEM: AN ASSESSMENT AND PROPOSED POLICIES 292 (1977). In analyzing the geographical impact of federal expenditures, it is important to distinguish between aid that is linked to particular locations and so generates differentials in land rents, and aid that goes to people without regard to their location and so does not create any rent differential. The list of federal aid programs that have geographical impact is quite long. Obvious programs are general revenue sharing (State and Local Fiscal Assistance Act, 31 U.S.C. §§ 1221-65) and Community Development block grants (Housing and Community Development Act of 1974, 42 U.S.C.

other federal programs have geographic impacts as well, and the net effect is difficult to estimate. Overall, federal expenditure programs may have shaped the current distribution of city sizes to a considerable degree. If the geographic impact of federal policies was understood, then it might be appropriate to allocate funds in a way that encouraged or discouraged certain city growth, or at least a passive policy toward city size might seek geographic neutrality.[37]

III
SUMMARY

The notion of an optimal distribution of city sizes seems to be more than a will-o'-the-wisp. The historical development of cities, the strong relationship between city size, quality of life, and worker productivity seem to suggest that city size is an important determinant of human welfare. Moreover, externalities are very important in urban areas, in traffic congestion and air pollution, for example. Therefore, unaided, decentralized markets seem unlikely to yield an optimal pattern of city sizes and efforts to deal with externalities directly may fall short of creating an optimal allocation of resources; therefore, a policy toward city size may be appropriate as a second best policy. While our understanding of the relationship between city size and human welfare is too primitive to justify active policies to promote a particular pattern of city sizes, the evidence is strong enough to urge the adoption of geographic neutrality among existing federal aid programs as a conscious goal, albeit one for which implementation techniques are uncertain. While such a passive policy does not deal precisely with the fundamental problems of the externalities, it is an inexpensive and administratively simple approach to reducing our loss of welfare from traffic congestion, crime, polluted air, and too few agglomeration economies.

---

§§ 5301-17) where urbanization enters the distribution formulae explicitly, and urban mass transportation capital grants (Urban Mass Transportation Assistance Act of 1970, 49 U.S.C. §§ 1601-130) where a few large cities receive disproportionate shares of the funds. Less obvious geographic impacts are found in the Comprehensive Employment and Training Act of 1973 (29 U.S.C. 801-992) emergency jobs program, where funds are disbursed only in areas where unemployment rates exceed seven percent. While this is ostensibly a program aimed at people, in fact it has a geographic emphasis. In 448 federal programs giving financial aid to state and local government, there are 146 programs with formulas for distributing funds geographically. Each formula presumably reflects some historical political balance but the net effect of the distribution formulas and project distributions seems unlikely to reflect a coherent philosophy of desirable geographic impacts. ADVISORY COMMISSION ON INTERGOVERNMENTAL RELATIONS, IN BRIEF: THE INTERGOVERNMENTAL GRANT SYSTEM: AN ASSESSMENT & PROPOSED POLICIES 7-9 (1979).

37. Any policy designed to deal explicitly with city size is a second best measure. For example, congestion tolls not only affect city size, but promote efficient utilization of transportation systems, lowering total transportation costs. Pollution prices promote technical change in pollution abatement as well as modifications in city size. Explicit geographic tax differentials or explicit geographic targets for the distribution of federal aid are much less precise methods for dealing with externalities.