From Public Economics, by William Vickrey
Edited by Richard Arnott, Kenneth Arrow, Anthony
Atkinson, and Jacques H. Dreze
Cambridge University Press, 1994.

17

The City as a Firm

Introduction

The purpose of this paper is to develop a rigorously optimizable model of a city, incorporating in this model two elements that seem to be vital to the development of large modern cities: transportation costs and economies of scale in the activities carried on within the city. If there were no transportation costs, so that goods and persons could be moved anywhere instantly and at no cost, and if there were no positive neighborhood effects, then activity could be indiscriminately scattered over the landscape instead of being concentrated in large urban complexes. Again, if there were no economies of scale, so that each activity could be carried on efficiently on as minute a scale as would be desired, activity could be organized in hamlets or small towns, each with a suitable complex of activities carried out on a small scale, with no need for the costly infrastructure characteristic of our large cities. Some degree of concentration into small cities might be accounted for by positive neighborhood effects of the kind that involves direct influence not mediated through a market transaction. The forces that bring our large cities into being are, however, predominantly those mediated through the purchase and sale of goods and services, so that the important positive externalities are of the market type, involving, in general, substantial economies of scale in at least some of the activities involved. Again, while some cities owe their location to an economically important natural feature of their site, such as a harbor or a mineral deposit, the size of the city that would coalesce around such a natural feature

The Economics of Public Services, Martin S. Feldstein and Robert F. Inman, Eds., Proceedings of a conference held by the International Economic Association at Turin (London: Macmillan; New York: Wiley, 1977), pp. 334–43.

would, in the absence of economies of scale, be far smaller than those we actually observe.

The existence of economies of scale within the activities involved immediately precludes an analysis based on perfect competition. On the other hand none of the manifold models of imperfect competition readily lends itself to incorporation in a precise model; and all of these models lack the optimizing properties that are being sought here. We shall, therefore, examine the problem of the optimal organization of a city in terms of some form of overall management structure, in effect regarding the city as a firm attempting to maximize some objective function. Even so, the inherent complexity of the situation dictates that on the one hand we keep the model stripped down to its barest essentials, and on the other that even then we stop somewhat short of full rigor of analysis, relying to a considerable extent on heuristics derived from classical models of perfect competition, or, perhaps, of a type of decentralized management derived from the analysis of Walras, Lange, and Lerner.

Littoral City

To construct a model stripped of all but the barest essentials while still retaining the essential elements of transportation cost and economies of scale, we consider a city laid out on a strip of land of uniform width along a shore. This strip will be considered to be long enough to contain all the activities to be carried out in the city, while leaving a vacant margin at either end, yet not long enough to enable a second city to be established on the adjacent portion of the strip. The sections of the strip not occupied by the city have no alternative use, while the land, if any, behind the strip is unusable for any purpose. Imports from, and exports to, other cities can be landed or loaded indifferently at any point along the strip at a uniform, delivered, or free-on-board price, while coastal ("intracity") transportation is handled at a constant cost per mile per unit of output of the ith commodity, t_i.

The character of the city is determined by the selection of a set of n activities to be carried on within the city, designated $i = 1, \ldots, n$. The level at which each activity is carried on is defined in terms of the volume of its single output, q_i. Economies of scale are represented by the fact that each activity requires a fixed quantum of imports f_i, measured in terms of aggregate delivered cost, in addition to a variable quantum of imports $m_i q_i$ proportional to the output q_i. The fixed element f_i is assumed to be large enough to assure that it will be nonoptimal to establish a second unit of activity i within the city. The extra fixed cost of setting up a second establishment would outweigh any saving that might be made in transportation costs within the city if a second and, possibly, closer source of supply for some users were provided. Each activity further requires the occupancy of $a_i q_i$ units of land, which, if the width of the strip is taken as unity, also implies the occupancy of a frontage equal to $a_i q_i$. In addition, each activity i may require as inputs amounts of the outputs of other activities, j, given by $b_{ij} q_i$. Inputs must be transported to, and outputs must be available at, the midpoint of the stretch of coastline occupied by each activity. The complication of the transportation cost of the inputs to the transportation process is avoided by stipulating that all of these inputs be imported. Finally, some or all of the output of an activity q_i may be exported at a net realization of $x_i p_i$, x_i being the amount exported and p_i the net price.

The optimization problem may be defined as one of maximizing

$$S = \sum p_i x_i - \sum f_i - \sum m_i q_i - T \qquad . (1)$$

where

$$T = \sum_i \sum_j (t_i \, d_{ij} \, b_{ij} \, q_j) \qquad (2)$$

is the total transportation cost, and

$$d_{ij} = (a_i q_i + a_j q_j)/2 + \sum_{s_i \leq s_k \leq s_j} (a_k q_k) \qquad d_{ii} = 0 \qquad (3)$$

is the distance between the centre of activity i and the centre of activity j, with s_i representing the sequence order of the ith activity along the strip, subject to

$$x_i + \sum_j b_{ij} q_j \leq q_i \qquad q_i > 0, \; x_j \geq 0 \qquad (4)$$

$s_1, s_2, \ldots, s_n$ is a permutation of $1, 2, \ldots, n$.

Mathematically, the solution of this problem for any given permutation of the activities along the strip is a fairly straightforward quadratic programming problem. Determining which of the permutations is the optimal one presents more of a problem, although, since the number of permutations is finite, the problem is, in mathematical principle, solvable, even though carrying out a solution by examining all permutations would very rapidly become infeasible for any large number of activities. The problem of developing an algorithm that would shorten the process is an interesting one, but is not the main concern here. In any case, sufficiently close results could probably be obtained by testing interchanges of each pair of adjacent activities, performing the interchange where an improvement is noted, and proceeding until further interchanges show no gain. If, when this stage is reached, further tests are made for the interchange of any two nonadjacent activities, and then, possibly for the removal of an activity from one place in the order to another, one might be fairly confident of being close to the optimum; but, unless some rather difficult theorem to this effect can be demonstrated, one could not be absolutely sure of a full optimum.

Decentralized Solution

For the present, more interest attaches to a solution by *tâtonnements* along the lines of a Lange–Lerner decentralized socialist state, in which each activity is in the hands of a manager instructed to behave in a pseudo-competitive manner, and in which the role of the city manager is limited to the determination of the sequence in which the activities will be located and of the rents that the activities will be required to pay for the space they occupy. Rents will be set equal to the marginal social cost of land occupancy, which in this case consists of the increase in transportation costs involved in the occupancy of more land. Indeed, if a given activity manager wishes to occupy more land, the consequence will be that all traffic that formerly traversed his frontage will now have to go that much further, assuming that, in effect, all facilities used by an activity can be freely shifted along the strip. The land rent of each activity will then be equal to the cost of the transportation carried out in front of the area that it occupies, and total transportation costs will be equal to total land rents. (This compares with the more familiar result for the two-dimensional case that transportation costs are twice land rents.) Determination of the proper sequence for the activities is a less automatic process, but we shall assume that by some means it is satisfactorily accomplished. Starting from some arbitrary set of prices, rents, land-occupancy levels, and output levels, each manager will determine his own marginal cost on the basis of these prices, rents, and the costs of transportation over the distances indicated by the intervening levels of land occupancy; add to his output for export if this cost is below the world price; determine his demand for the outputs of other activities; and report the results. These are fed into a second round, in which each manager aggregates the demands made by other managers, so as to determine his own level of output for domestic consumption; adds, where appropriate, an amount for export, to get his new level of output; and recomputes his marginal cost. This is repeated until an equilibrium is reached in which the price of each output is equal to its marginal cost, and is either equal to the world price (where the item is being exported) or equal to or greater than the world price (where it is not being exported). The city manager can then test whether an interchange is desirable by examining the total per-mile freight costs of the items of input into two adjacent activities, i and j. If the costs for the inputs that are brought to i past j and to j past i are more than half of the total per-mile freight costs of all inputs to

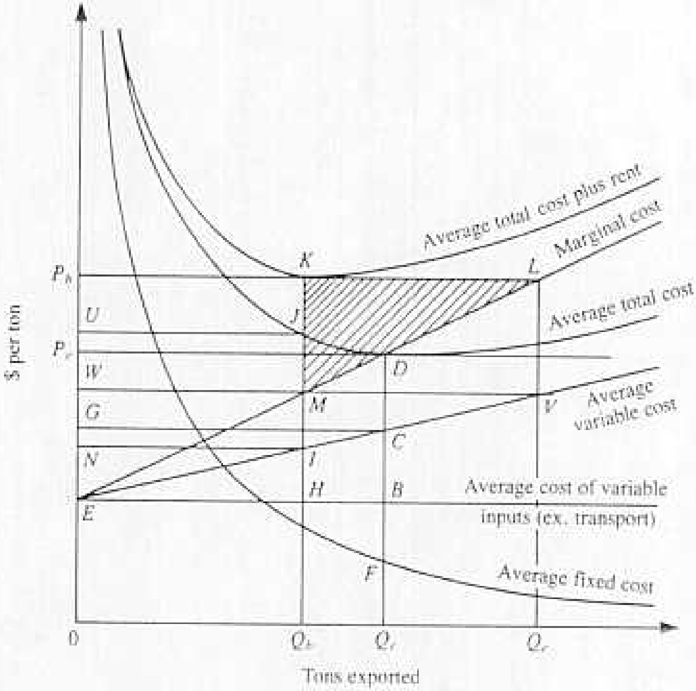

Figure 17.1

these activities (other than between i and j), the exchange should be made, but otherwise it should not.

Competition Between Cities

Although not all cities are of the same type (i.e., carry on the same mix of activities), it would seem reasonable to assume that there is a sufficiently large number of cities of each type, equally advantageously located with respect to world markets, and with no lack of additional vacant strips of land on which other cities might be established, to allow us to subject the city to the analysis developed for the firm in perfect competition. If, for simplicity, we assume that each city type exports only one commodity, we can represent the situation as in figure 17.1 in which the volume of the exported commodity is on the horizontal axis and costs per unit are measured vertically. The total average cost is made up of three components, corresponding to the last three terms of equation (1). We have, first, the constant variable cost other than transportation, represented by $\Sigma\, m_i q_i$ in the equation and shown as $0E$ in figure 17.1, and the horizontal line EHB.

The transportation-cost term, T, is homogeneous of second degree in the q_i, so that the average transportation cost is proportional to the level of activity and to the level of export x_i, and can be represented by the difference between the straight lines EHB and $EICV$. $EICV$ is then the average variable cost including transportation. The corresponding marginal cost is $EMDL$, with $BD = CD$, and so on. The fixed-cost term yields the equilateral hyperbola through F as the average fixed cost. When added to the average variable cost, this yields the average total-cost curve JD, intersecting the marginal-cost curve at its minimum point, D.

Then, if we have an unlimited number of cities of each type, with free entry for new cities and perfect competition operating among the various cities, equilibrium should establish itself with the price of the exported commodity settling down to the level P_e relative to the cost structure, with price equal to marginal cost equal to average cost, from the standpoint of the city as a whole. However, from the standpoint of the individual firm within the city, the picture will look slightly different. The price that it will be quoting (calculated at marginal cost, other prices being considered fixed) will cover the delivered cost of the variable inputs plus the rental charged for the land it occupies, but will not cover the fixed costs, f_i. For this the firm will need a subsidy from the city manager. For any individual activity, there need not be a relation between the amount of land rent paid to the city manager and the amount of the subsidy needed to cover the fixed costs.

However, for the city as a whole, rents will just balance the subsidy, as can be seen from figure 17.1 where the transportation costs given by the rectangle $EBCG$ are equal to the fixed costs represented by $GCDU$, by virtue of the fact that $BC = CD$. But we have seen that the rents on each piece of land are equal to the marginal social opportunity cost represented by the transportation costs incurred in front of the property; hence, total land rents are equal to total transportation costs. In the aggregate, then, aggregate land rents at rates representing the marginal social cost of land occupancy will be just sufficient, and no more, to pay the subsidies required to permit the individual activities to lower their prices to marginal cost. Thus, to the proposition of Henry George that, as a matter of justice, the land rents that have arisen through no effort on the part of any individual land-owner should be appropriated for public purposes, and the suggestion of Harold Hotelling that taxes on land rents would be an appropriate means of financing the intramarginal cost residues of increasing-return industries pricing their output at marginal cost, we add the following theorem, which we call the "*GHV*" theorem.

> *In an economy of efficiently organized cities in a state of perfect competition with each other, the aggregate of the land rents (calculated as the marginal social cost of holding land) generated by the urban*

> *agglomeration produced by the existence of activities with economies of scale within the city will equal the subsidies required to enable these activities to sell their output at prices equal to their respective marginal costs.*

In other words, if a perfectly efficient world is to be organized on a classical decentralized basis, it is necessary that all of the land rent generated by the presence of the city be appropriated to the subsidy of the decreasing-cost industries within the city. If any of these land rents are appropriated by private landlords for their own purposes, this action will preclude the achievement of complete efficiency.

The Scope of the GHV Theorem

Thus far, this theorem has been illustrated rather than proved, but there is reason to believe that it is valid well beyond the rather oversimplified example given above. For example, it would seem that going from a one-dimensional to a two-dimensional, or even a three-dimensional, model would not destroy the validity of the theorem. In a two-dimensional model, transportation costs per unit of output would tend to vary as the square root of the volume of output, and marginal transportation costs would then be 3/2 of average transportation costs. In terms of figure 17.1, we would replace the line *EIC* with a parabola with axis *EB*, and *EDL* would be another parabola, with $BD = 3/2\ BC$. At the minimum-cost point *D*, the fixed cost would thus be equal to half the transportation cost. There are also a number of models that, for a two-dimensional case, produce an equality between transportation cost and twice the total land rents. Thus, again, we have land rents equal to the fixed costs. If we allow for the possibility of expanding vertically, say by thinking of the city as being, in effect, constructed in a k-dimensional space, with $2 < k < 3$, then we can have $T = kR = kF$, with, again, $R = F$, though in this case T may have to be thought of as including the cost of lifts, stairs, and the like.

Again, the assumption of a linear production function with a constant term does not appear to be critical, since any production function can be approximated, in the neighborhood of the optimal point, by a linear function tangent to the actual function at the equilibrium point, with a constant term that can be taken as the effective intramarginal residue. Similarly, the fact that transportation costs do not vary simply in proportion to distance, but vary in a curvilinear fashion, may be allowed for by replacing the actual transportation-cost function with one that is a linear function tangent to the actual one at the equilibrium point, and by taking the element of cost represented by the intercept and incorporating it, as packing or loading costs, in the costs of the shipping activity. Among the matters that are not so simple to deal with, but that probably can be handled along analogous lines, is the possibility that part of the fixed costs of activities may take the form of a fixed land requirement, or that inputs to the transportation service are locally produced.

An element that may be more likely to disable the theorem is the fact that, in practice, much transportation does not take the form of a direct origin-to-destination shipment, but is multi-purpose in character, as when an individual stops at the barber on his way to work, or, more importantly, when a delivery truck drops off or picks up shipments at various points along a route. Computing a relevant marginal cost under these circumstances is difficult, and it is not even clear that there will be any simple algorithm for determining the optimum pattern on a decentralized basis (consider the travelling-salesman problem).

Another caveat to be kept in mind is that the theorem applies only to the intramarginal residues occurring in those industries that are either entirely local in their markets or are replicated in similar cities. A local newspaper or a television station, for example, is a factor in the agglomerative power of a given city, and the land rents generated by this agglomerative power will cover the intramarginal costs of these activities. On the other hand whereas publication in a given city of a unique book with a worldwide distribution may add, in some degree, to the agglomerative power of the city, it is unlikely to do so to a degree that would cause land rents to rise sufficiently to cover the writing, typesetting and preparation costs involved. In effect, the theorem applies to local economies of scale, not to global ones. However, the economies of scale involved are not necessarily manifested in large organizations: the factory that makes buttons of a given model more cheaply in lots of 600 than in lots of 300 is operating under conditions of economies of scale, even though there would be no reduction in average cost if it produced 500 buttons in each of 200, instead of 100, models.

Optimizing in an Environment of Nonoptimizing Cities

A more immediate question, given the remoteness of a fully optimized situation, is how a city might optimize in a context in which other cities are behaving in a nonoptimal manner, and what the consequences of such action might be. This involves, first, defining a specific nonoptimal type of behavior that is to be ascribed to the outer universe of cities. Given the possibilities for various varieties of imperfect competition inherent in the prevalence of economies of scale, this is a wide-open opportunity for variation. Since we are concerned with the relation of land rents to city efficiency, let us assume that the one constraint against optimality is that land rents, instead of being used to subsidize the decreasing-cost activities so as to permit marginal-cost pricing, are used by landlords for their own private purposes (for instance, consumption of additional imports). This will mean that individual activities will have to be financed from their own receipts without subsidy, and that prices must cover the full average cost, inclusive of purchased inputs, land rents, and fixed costs. Let us assume, further, that land rents are optimally set so as to reflect the marginal cost of land occupancy, as determined by the cost of transportation. The average variable cost will then be $0E$ plus twice the transportation cost, or $EMDL$, and the average total cost, inclusive of rent, will be indicated by the curve through K. We may then suppose the world price relationships to be set at the level represented by P_b, the output being at the level represented by Q_b. Of the total proceeds $0P_bKQ_b$, $0ENQ_b$ represents variable direct imports, $NIHE$ the imports for transportation, $NIMW$ is the rents consumed by landlords, and WP_bKM the fixed costs.

If, now, in the context of such inefficient cities, a given city wishes to optimize its own operations, it should operate at the point where marginal cost, net, equals the world price, or at L. Expansion from K to L produces additional revenues of KLQ_bQ_e against additional costs of MLQ_eQ_b, leaving a net gain of MKL. This gain, in the linear case, is equal to the former level of rents, with $Q_e = 2Q_b = 1.414Q_c$. For the k-dimensional case, the corresponding results are

$$Q_c = \left(\frac{kF}{t}\right)^{k/(k+1)}$$

where t is the transportation cost involved in a city producing one unit of output

$$Q_b = Q_c\left(\frac{k}{k+1}\right)^{k/(k+1)}$$

and

$$Q_e = \left(\frac{k+1}{k}\right)^k Q_b = Q_c\left(\frac{1+k}{k+1}\right)^{k^2/(k+1)}.$$

The land rents to be collected at K will be

$$L_b = F\frac{k}{k+1}$$

and the total surplus available at L will be

$$S_e = F\left[\left(\frac{k+1}{k}\right)^k - 1\right].$$

At L, however, land rents at the social opportunity cost would be

$$F\left(\frac{k+1}{k}\right)^{k}$$

so that the surplus is insufficient, by the amount F of the fixed costs, to cover these rents.

In effect, if the landlords constitute the ruling class of the city, it would be in their own long-run interests, with world prices at P_b, to levy against their rents to subsidize to the extent of F, inducing in the long run (assuming that no other cities do likewise) an expansion to L and an increase in the gross land rents by more than the subsidy F. Of course, this process would take time. Indeed, at the initial point K, land rents are less than F, and would be insufficient at that point to do the whole job. It is only as the city expands beyond the output Q_c that land rents become sufficient to cover the entire intramarginal costs of the decreasing-cost industries.

Surplus and Efficiency

Of course, if all cities started to do this simultaneously, one might ask in what sense a shift from K to D is an increase in efficiency, since it eliminates surplus. The answer is to be found in what happens to the returns available to whatever are, ultimately, the scarce resources in the universe under discussion. An overall shortage of suitable sites for cities would lead, of course, to a surplus rent. A more interesting concept would be that of a limited labor force, freely mobile among cities. Expansion of the scale of cities, and, possibly, the establishment of new cities, would lead to an increase in the demand for labor, and an increase in its wages, represented in this model by the level of the input coefficients to the labor supply activity, representing an increase in the supply of consumer goods necessary to produce a unit of labor for use in the other activities in the city. Full competition among cities would, in this instance, lead (a) to the elimination of any new land rents to landlords, over and above the amounts taken to finance the fixed costs; and (b) to the transfer of this surplus, plus the additional surplus generated from the increase in efficiency to labor, at least in the case where there is no shortage of city sites and, hence, conceptually free entry for additional cities.

Summary

While it is dangerous to extrapolate too freely from admittedly oversimplified and even caricatured models (such as the above) to the complexities of the real world, it seems not too rash to draw the conclusion that urban land rents are, fundamentally, a reflection of the economies of scale of the activities that are carried on within the city, and that efficient organization of a city, or even of the urban life of a nation as a whole, requires that these land rents, or their equivalent, be devoted primarily to the financing of the intramarginal residues that represent the difference between revenues derived from prices set at marginal cost and the total cost of the activities characterized by increasing returns. This means that revenues derived from these rents must defray not only those overhead costs of government and public services that are not marginally attributable to specific outputs, but also the intramarginal elements of most public utility costs, as well as those of a considerable range of activities ordinarily thought of as belonging to the private sector. Such a subsidy is especially crucial for mass transit and for other forms of transportation where economies of scale are significant, given the special relationship existing between transportation and land values.

Use of land rents, or, at least, of a major fraction of them, for public purposes is therefore not merely an ethical imperative, derived from categorization of these rents as an unearned income derived from private appropriation of publicly created values, but is, even more importantly, a fundamental requirement for economic efficiency. Cities that take the lead in such public use of land rents may find that in the long run the subsidy is self-financing through the enhancement in land values that results, while cities that lag may find that in the long run they are unable to compete in national or world markets with the cities that manage to organize themselves more efficiently through the use of land rents to cover intramarginal residues. There will, of course, be many an agonizing slip between abstract economic analysis and cold political and economic reality. Lack of comprehension, political intervention, strategic recalcitrance, and the inertia associated with heavy commitments of fixed capital in what the French so aptly term *immeubles* may slow the processes involved to a glacial pace. But the fundamental tendencies and requirements inherent in the very nature of the city can be ignored only at great peril to its economic health.