Urban Economics and Public Policy

Second Edition

JAMES HEILBRUN
FORDHAM UNIVERSITY

ST. MARTIN'S PRESS NEW YORK

For Carol

ACKNOWLEDGMENTS

p. 73, Figure 4.1: Adapted from figure 6, p. 86, in William Alonso, "Location Theory," reprinted from *Regional Development and Planning* by John Friedman and William Alonso, eds., by permission of the MIT Press, Cambridge, Mass. © 1964 by the Massachusetts Institute of Technology.

p. 115, Table 5.5: Changes in the hierarchy of trade centers in Saskatchewan, 1941–61. Data from Gerald Hodge, "The Prediction of Trade Center Viability in the Great Plains," *Papers* of the Regional Science Association, vol. xv, table 2, p. 95.

p. 138, Figure 6.5: Adapted from William Alonso, *Location and Land Use* (Cambridge, Mass.: Harvard University Press, 1964), figure 32, p. 112.

Library of Congress Catalog Card Number: 80–52365
Copyright © 1981 by St. Martin's Press, Inc.
Manufactured in the United States of America.
54321
fedcba
For information, write St. Martin's Press, Inc.,
175 Fifth Avenue, New York, N.Y. 10010

Cover design: Quentin Fiore

ISBN: 0–312–83441–1

orders increased from 90 in 1941 to 125 in 1961, while the number in the three lowest orders fell from 816 to 654. The three lowest orders made up 90 percent of all places in 1941, but only 84 percent in 1961.[21] Studies by Berry and others of central place systems in farming areas of the U.S. Midwest reveal similar tendencies there.[22] These patterns of development are quite similar to the one suggested by the aggregate data for the United States in Table 5.4: under the impact of economic change and growth, central service functions have been gradually concentrating into fewer and larger places.

INTRAMETROPOLITAN PATTERNS ARE NOT EXPLAINED BY CENTRAL PLACE THEORY

It is no accident that this discussion of central place theory has focused on villages, towns, and small cities rather than on the great urban centers or metropolitan areas of the nation. True, we can observe hierarchical differences among large cities as well as among smaller places. Major cities and metropolitan areas can be differentiated into hierarchical classes by the order of services they provide and the regions for which they are the dominant provider. Thus they, as well as the smallest towns, are a part of the "system of cities," and our insight into urban phenomena is greatly enhanced by understanding that. Yet within metropolitan areas themselves it is not easy to discern the hierarchical pattern of market areas (and of centers) for the distribution of goods and services that is described in classical central place theory. Population is so dense and travel relatively so easy that shoppers habitually visit many suppliers, and the market areas of suppliers cease to be even approximately exclusive. Under these conditions, as Berry has pointed out, the locational pattern of central service firms becomes a complex of "ribbons" and "specialized areas" as well as of "centers."[23]

In the face of this complexity economists have, after the manner of their kind, looked for some underlying, simplifying principle that would explain the essence of the whole intrametropolitan pattern of location. They have employed as their point of departure a model that does not break the metropolis up into a hierarchically related system of submarkets but assumes instead that it is a single market area, organized into specialized districts around a single, unchallenged center. It is to this model that we turn in the next chapter.

---

[21] Hodge, table 2, p. 95.

[22] See Berry, *Geography of Market Centers*, pp. 114–17, and references cited therein.

[23] Ibid., pp. 44–58 and 117–24.

Site Rent, Land-Use Patterns, and the Form of the City

SIX

Just as transport costs influence the location of producers and therefore of cities, so, too, they systematically affect locational patterns within the city itself. We are all conscious of the general form of the city: concentrated activity and development at the center; a gradual decline of intensity as one moves out toward the edge. This characteristic form is constantly impressed on us by the clusters of tall buildings and the dense crowds of people and vehicles we see at the center, so dramatically different from the lower skyline and the smaller crowds we find at the periphery.

It is not difficult to construct economic models that will generate this easily observed pattern as a function of transport cost and the need for accessibility. Although these relatively simple models do not account for the full complexity of land-use arrangements that are found in the city, they do explain successfully the general pattern of land use, land value, and density of development. In addition, they are vitally important because they illuminate the process by which the real estate market sorts out potential occupants of land, allowing those who can make the most productive use of central sites to obtain them and pushing those less dependent on centrality out toward the edge. From an understanding of this process much can be gained. The city planner learns the strengths and weaknesses of a free market in land as a rational allocator of scarce central sites. Students of housing and urban renewal gain insight into the economics of "land-use succession," the process by which "renewal" occurs (or fails to occur) in a freely functioning market. The transportation analyst learns the interdependence of land values, land use, and the transportation network.

In all such models of land use, the price of urban land at various sites plays a crucial role. This price can be expressed either as the capital value of a site or as its annual rental value. Although land is more commonly sold than rented in the United States, it is usually more convenient for purposes of analysis to use annual rent rather than capital value as the measure of price. The relationship between the two concepts is, in any case, simple enough: the capital or market value of a site is the present value of the stream of net returns it is expected to yield in the future. Since land in uses other than agriculture does not "wear out," its expected future life is infinitely long. If, then, its yield is expected to be constant per year, the expression for its present capital value reduces to the formula for evaluating the worth of a perpetual income: capital value equals the expected perpetual annual net return divided by the interest rate appropriate for capitalizing an income of that degree of risk.

A SIMPLE MODEL: RESIDENTIAL LAND USE

Let us start with a drastically simplified model based on the following assumptions: (1) a city has sprung up on a flat plain, or transport surface, of the sort assumed in the explanation of central place theory in the preceding chapter; (2) all production and distribution activity in this city takes place at a single point at its center; (3) the populace consists of families of uniform size, taste, and income who must live in rented single family homes of uniform house and lot dimensions ranged around the central production and distribution point; (4) the cost of building and maintaining houses (site rent excluded) is constant throughout the city.[1] Since housing costs other than site rent are spatially invariant, differences in total rent paid by tenants at two different sites clearly represent differences in the site-rent component of the total. "Site rent" itself we define as the rent paid for a site less the rent it could command in an agricultural use.

Under these assumptions it can be shown that site rent for residential lots will be highest adjacent to the business center, will decline along any radial from that point, and will fall to zero at the periphery. Beyond the periphery land will command agricultural rent only.

FIGURE 6.1(a)

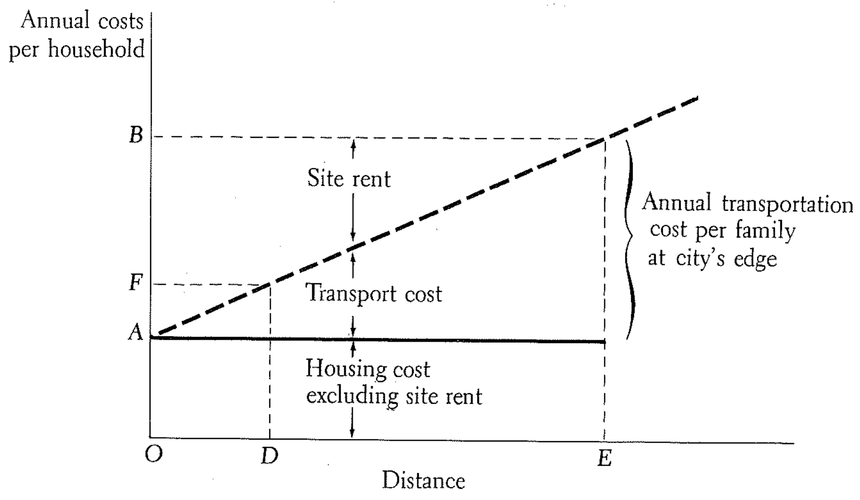
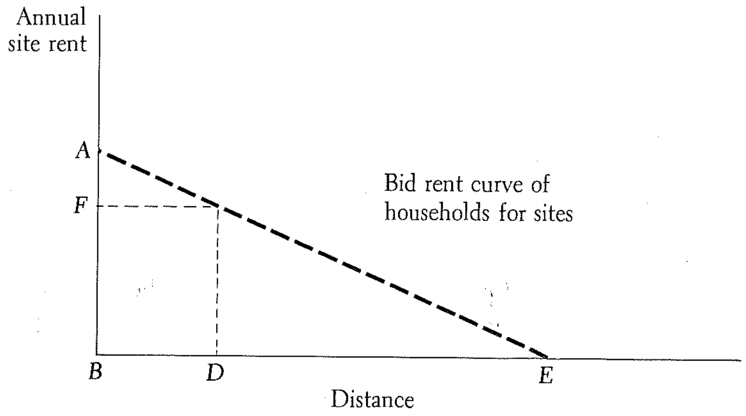


FIGURE 6.1(b)



In Figure 6.1(a) we place the center of the city at O and measure distance from the center along the horizontal axis to the right. The vertical

---

[1] The land-use model that we here apply to the city had its origin in J. H. von Thünen's work *Der isolierte Staat*, published in 1826. Von Thünen made use of the featureless plain and the need for access to a market center to explain the pattern of agricultural land uses that typically formed around a market town. In the twentieth century, urban economists have relied almost exclusively on a similar model to explain the pattern of urban land use. For a summary of twentieth-century theories of the economics of urban land use down to 1960, see William Alonso, *Location and Land Use* (Cambridge, Mass.: Harvard University Press, 1964), ch. 1. The heart of Alonso's book is a model, far more elaborate than the one attempted here, that comprehends residential, business, and agricultural land uses within the now familiar von Thünen framework.

axis measures costs per family unit on an annual basis. Let us assume that for a given population and lot size and with no vacancies, the area needed to accommodate all families is provided when the city extends as far out as point *E* on the horizontal axis.

The vertical distance *OA* on the cost axis represents the annual cost of housing per unit, exclusive of site rent. This cost is equal for all locations, as indicated by the horizontal line at height *OA*. Since jobs and all consumer services are concentrated at the center, each household must bear an annual transportation bill for trips to the center and back. The annual cost of these trips increases with the household's distance from O and is shown by the rising transportation cost curve. Transport costs per family rise from zero at the city center to a maximum of *AB* for households located at the city's edge.

We can now show how site rent would arise as a payment for the saving in transport cost that could be obtained by living at any particular site. Consider, for example, a site adjacent to the city center. The owner of such a parcel could ask an occupant to pay as much as *AB* in site rent. If all houses are occupied, a householder would be willing to pay that much site rent to locate at the center, since the alternative would be to move just beyond the present edge of the city, where site rent would be zero but it would be necessary to bear transportation costs equal to *AB*.

Under the usual assumptions about competition of occupants for houses and of landowners for maximum rent, the same argument can be extended to all sites: at any point within the city, occupants would be willing to pay as site rent the difference between transportation costs at that site and the higher transportation costs that would be incurred if they moved to the city's edge. Under these conditions the combined cost of housing plus transportation plus site rent is the same at all sites and equals *OB* in Figure 6.1(a).

In Figure 6.1(b) site rent itself is plotted against distance from the center. (We have simply inverted the site-rent triangle of Figure 6.1(a), while retaining its dimensions and labels.) Rent declines from *BA* at the center to zero at the edge. Thus our simple model approximates observed reality, in which land values are highest downtown and fall off with some consistency as one moves out to the metropolitan periphery.

The declining line in Figure 6.1(b) is usually described as a "bid-rent curve," since it shows the maximum site rent that households would willingly bid at each location. Since by assumption in this model no other land users compete with households for urban sites, the bid-rent curve of households becomes in fact the "rent gradient" for the city. Later, when introducing other land uses, we will show that the actual rent gradient in a city is produced from competition among uses and can be depicted by graphically combining their several bid-rent curves.

The Economic Character of Site Rent

The very simplicity of the model employed so far is a virtue, since it enables us to see clearly many of the important characteristics of rent in general and of site rent in particular. "Economic rent" is defined by economists as a payment to a factor of production in excess of its opportunity cost (for which reason it is also called an "economic surplus"). The opportunity cost of a factor in a given use is the payment it could command in its next best employment. The next best employment for urban sites is as agricultural land. Since site rent is a payment in excess of this opportunity cost for urban sites, it clearly conforms with the general definition of economic rent.

Factors of production command an economic rent only to the extent that they are "scarce," and they are scarce in the long run only to the extent that they are nonreproducible. Again urban sites illustrate the general case. While land on which to build is plentiful, land with accessibility to economic centers is scarce; it is the scarce attribute of accessibility that gives rise to site rent. In our simplified model sites can be added indefinitely by extending the edges of the city, but each incremental ring of sites has less accessibility than the adjacent ring closer to the center and commands correspondingly less rent. Accessibility to a given center cannot be reproduced, though it can be altered by changes in the technology of transportation.

Although rent is a payment in excess of opportunity cost and can therefore be described as an economic surplus, it is nevertheless a payment equal in value to the margainal product of the factor to which it accrues. Site-rent payments are therefore consistent in every respect with the marginal productivity theory of distribution and, as we will see, perform an essential function in bringing about an efficient allocation of land among competing uses. That site rent equals the marginal product of the site can easily be seen from Figure 6.1(a). Compare, for example, site *D* with site *E*. The rent at *E* is zero; at *D* rent equals *FB*. Suppose the site at *D* were vacant and a family moved there from *E*. At *E* the family would have borne transportation costs equal to *AB*, as compared with transportation costs of *AF* at *D*. The reduction in transportation costs brought about by the move is given by $AB - AF = FB$. This reduction is a saving in real resource costs, and this saving is precisely the marginal product of the occupied site at *D*. It is also precisely the competitive site rent payable at *D*. Hence under competitive conditions rent equals marginal product.

Limitations of the Simple Model

The model developed above yields results that are unrealistic in several respects. First of all, numerous studies have been made in recent years of actual urban site-rent gradients. Most of these have concluded that the gradients are not linear—that is, they do not have a constant slope as in Figure 6.1(b). Instead they tend to be steepest at the center and to flatten out toward the edge, a shape that is often best approximated by a negative exponential curve.[2] Such a curve has the characteristic of declining at a constant *relative* rate, instead of at the constant *absolute* rate of a linear gradient. Along a negative exponential curve, site rent would decline by the same percentage for each mile of movement away from the center. It should be added that the rate of decline of the rent gradient appears to vary widely from city to city.

Our initial site-rent model is unrealistic in a second respect. We know from casual observation that density of urban settlement is not spatially uniform, as the model requires, but instead is much greater at the center. Again, statistical studies have verified the pattern. Like the gradient of site rent, the gradient of population density in modern cities has usually been best approximated by a negative exponential curve.[3]

Finally, greater population density at the center is associated with greater density of improvements—that is, with more cubic feet of building per acre. The visible evidence of this is, of course, the higher skyline we observe at the center, not just for office buildings, but for apartment structures as well.

Figure 6.2 shows the gradient of land value per square foot on the west side of Manhattan. The curve traces values along a ray extending from the edge of the central business district at Sixty-second Street to the northern tip of the island. It displays the diminishing slope with movement away from the center that is typical of large modern cities. In the next section we show how such a gradient can be generated by our residential model if we relax some of its initial simplifying assumptions.

A MODEL THAT GENERATES SYSTEMATIC VARIATION IN DENSITY OF DEVELOPMENT

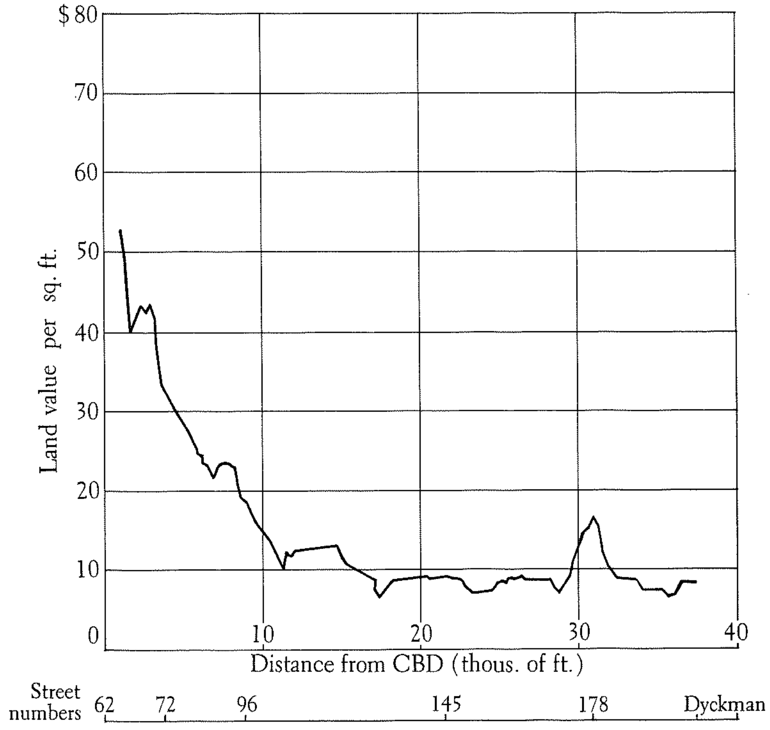
In constructing a more realistic model, we retain the assumption of a mononuclear city built on a transport surface with all commercial activity

---

[2] See Edwin S. Mills, "The Value of Urban Land," in Harvey S. Perloff, ed., *The Quality of the Urban Environment* (Washington, D.C.: Resources for the Future, 1969). In addition to reporting his own findings (which are discussed at length later in this chapter), Mills summarizes the work of several other investigators.

[3] See, for example, Colin Clark, *Population Growth and Land Use* (New York: St. Martin's Press, 1969), ch. 9; Richard Muth, *Cities and Housing* (Chicago: University of Chicago Press, 1969), ch. 7; Edwin S. Mills, "Urban Density Gradients," *Urban Studies*, February 1970, pp. 5–20; and Brian J. L. Berry and Frank E. Horton, *Geographic Perspectives on Urban Systems* (Englewood Cliffs, N.J.: Prentice-Hall, 1970), ch. 9.

FIGURE 6.2
Gradient of Land Value on the West Side of Manhattan Island, 1970[a]



[a] Five-block moving average of land value per square foot

Source: "The Empirical Investigation of a Residential Land Value Model," unpublished doctoral dissertation by Joseph E. Earley, Fordham University, 1974.

concentrated at a point in the center. Under these conditions, the transport-cost gradient remains as in Figure 6.1(a)—a linear function, rising as distance from the center increases. We continue to assume that families are uniform in size, taste, and income and that they occupy units of standard size. However, the requirement of uniform density of housing per acre is abandoned. Instead, we allow developers to pile units up on a given lot by building vertically. It is assumed that building developers do not own the sites on which they build but lease them from site owners. Tenants, in turn, rent space in buildings from the developers. Thus the market has three tiers: site owners, building owners, tenants. This institutional arrangement is used only occasionally in the United States, but it

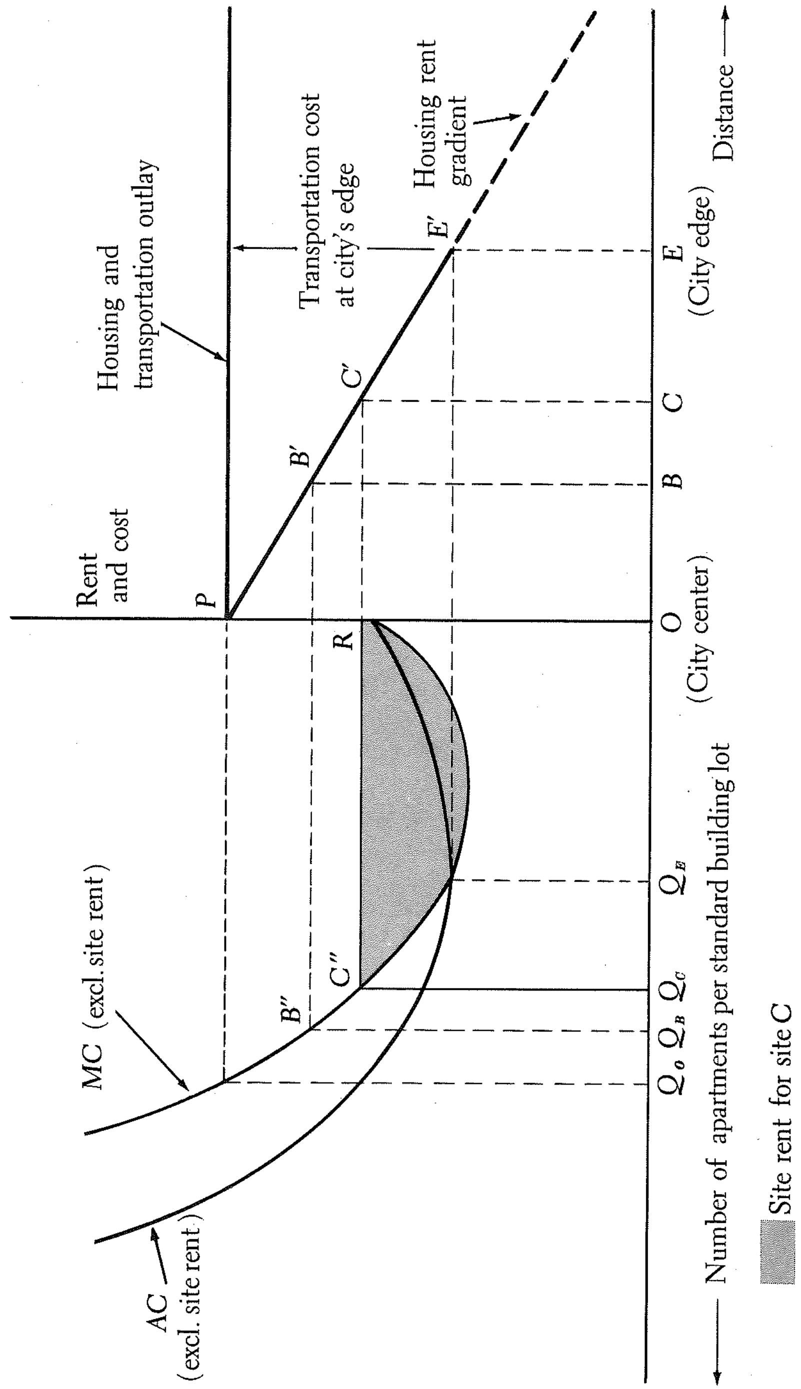


FIGURE 6.3

is quite common in England. It is assumed here only as a convenience, and does not affect the outcome of the analysis.

The new model is depicted in Figure 6.3. The right-hand panel shows transport costs and housing rent per family at various distances from the city center at O. The left-hand panel shows annualized costs to developers of building and operating standard size apartment units in multiple dwellings of varying heights.

Let us first examine the right-hand panel. Rent per apartment and transport cost per family are measured on the vertical scale between the two panels. To simplify the argument, we assume that all families arrange their budgets so as to pay a constant sum (equal for all families) for the combination of housing plus transportation. This outlay equals *OP* along the vertical scale, and the horizontal line through *P* indicates that the outlay is constant for all families regardless of location. At the center, transportation cost per family is zero, so *OP* is available as rent per apartment. Since transportation costs per family increase with mileage from the center, the sum available for rent decreases with distance and is shown by the housing-rent gradient, which declines from *P* at the center to *E′* at the city's edge. Transportation costs at any site away from the center are shown by the vertical distance between the horizontal line through *P* and the declining housing rent gradient.

The left-hand panel is a mirror image of the conventional diagram of the long-run cost curves of a business firm, in which *AC* represents average cost and *MC*, marginal cost. In this instance the curves measure annual long-run average and marginal cost, per apartment, of building and operating apartment structures of varying heights on lots of standard size. Cost includes normal profit but, unlike the conventional case, excludes site rent. The horizontal scale shows number of apartments per structure, increasing in the leftward direction. Moving leftward from the vertical axis at the center, there is an initial stage of declining average cost (i.e., *increasing* returns) as structure size increases from zero. This decline reflects the fact that construction and operating cost per square foot diminishes as floor area size increases, if other dimensions are held constant.[4] Thus a very small building on the standard lot would probably cost more per unit than a building that made maximum permissible use of the ground area. However, as more stories are added, the increasing cost of building higher eventually offsets this initial gain. Diminishing returns to the use of land set in for buildings larger than Q_E. Marginal and average costs per apartment rise, beyond a certain structure height, because the construction cost per apartment of adding additional stories

---

[4] See John J. Costonis, *Space Adrift* (Urbana, Ill.: University of Illinois Press, 1974), pp. 95 and 101–02.

begins to rise as the building grows taller. The rise in cost has two sources: first, the need for heavier foundations and structural elements and, second, the shrinkage of rentable interior space per story as height rises, on account of required building setbacks and the need for additional space-using elevator shafts.[5]

Taken together, the two panels of Figure 6.3 show the intensity of development that will occur and the site rent that will be generated at any distance out from the center. For example, at site *C*, tenants will be willing to pay CC' in housing rent. We assume that builder-operators behave as perfect competitors, treating the going rent level at any distance from the center as if it were independent of their own decision to build. Builders at distance *OC* from the center will be guided by rent level CC' in determining how intensively to develop their sites. They will build structures to the height at which the last story adds apartments whose marginal cost just equals the rent offer CC'. In other words, they follow the conventional rule for the perfect competitor, which is to extend output to the point at which marginal cost just equals price. Price at site *C* is shown by the horizontal line extending left from C'. This intersects the marginal cost curve at C''. The builder at *C* therefore puts up a structure containing Q_C apartments.

What about site rent, which is not included in the cost curves depicted in Figure 6.3? The maximum annual site rent that a builder could pay at any site is the difference between total annual housing rent and total annual cost (excluding site rent) for an optimum size structure at that distance from the center. Total housing rent that could be realized in a building at site *C* containing Q_C apartments equals the product of the rent per unit times the number of units, which is shown by the area of the rectangle $OQ_CC''R$. Total cost (excluding site rent) equals the sum of all marginal costs and is shown, for site *C*, by the area under the marginal cost curve up to the quantity Q_C. Maximum site rent is the difference between these two areas, shown by the saucer-shaped shaded area on the left-hand panel. Since builders are competing to obtain scarce sites, actual site rent will in the long run tend to equal the indicated maximum.

Our example illustrates the fact that site rent arises as a residual in the pricing process. Out of the proceeds from the sale of output (in this instance, housing services) producers must pay the going market price for whatever reproducible inputs they employ, including capital. The owners of nonreproducible scarce factors, such as urban sites, can then command as the price for the use of such factors whatever residual remains when all other inputs have been paid for by producers. The size of this

---

[5] Concerning these increasing costs see Ralph Turvey, *The Economics of Real Property* (London: Allen and Unwin, 1957), pp. 15–16; and Costonis, pp. 95–97 and 101–02.

residual depends on the price producers can obtain for their products and, therefore, on the level of demand. It is for this reason that rent is usually said to be "price determined" rather than "price determining."

We can now use Figure 6.3 to demonstrate two points. First, as we move toward the city center, say from site *C* to site *B*, the intensity of development increases. Builders at *B* can obtain higher rents per unit than builders at *C*. They therefore carry development on the site farther before reaching the point at which $MC =$ rent: Q_B is to the left of Q_C, indicating that buildings at *B* are taller than those at *C*.

Second, it is easily seen that site rent is higher at *B* than at *C*, since the saucer-shaped site-rent area grows larger as rent per apartment rises to BB' and the number of apartments increases from Q_C to Q_B. But we can say more: the increase in site rent per mile of movement toward the center is greater for each successive mile. This means that the bid-rent curve for housing sites that could be derived from the left-hand panel of Figure 6.3 does not have the constant slope displayed by the simpler model in Figure 6.1. Rather, the increased intensity of development as we move toward the center generates increasingly large increments to site rent, so the bid-rent curve for sites grows steadily steeper as it nears the center. This result can easily be deduced from the shape of the site-rent area. Each mile of movement toward the center in the right-hand panel generates a constant increase in housing rent per apartment equal to the transportation cost saved by locating one mile nearer the center; apartment rent rises along a gradient of constant slope. But these successive, equal-per-mile increments to apartment rent, which could be measured along the vertical center scale, generate *increasing* increments to site rent, as demonstrated by the fact that the layers added to the saucer-shaped site-rent area are successively longer as they pile up in the vertical direction. These layers are longer only because intensity of development increases as we move toward the city center.[6]

Intensity of development at the city's edge equals Q_E apartments per building lot. At that location, housing rent just covers average unit cost at the lowest level at which a developer could break even. Nothing is available for site rent, which is therefore zero. Tenants need not pay

---

[6] Edgar M. Hoover used a price line and a rising marginal cost function to derive a site-rent gradient of increasing slope for agricultural land uses in *Location Theory and the Shoe and Leather Industries* (Cambridge, Mass.: Harvard University Press, 1937), pp. 24–26. The model in fig. 6.3 applies his argument to the urban case. It is interesting to note, however, that the entire argument of this section (and the following one on competing land uses) could equally well have been derived from the work of Alfred Marshall. See his *Principles of Economics*, 8th ed. (London: Macmillan, 1930), pp. 447–50.

For a somewhat different derivation of an increasingly sloped urban site-rent gradient, see Harold Brodsky, "Residential Land and Improvement Values in a Central City," *Land Economics*, August 1970, pp. 229–47.

site rent to live in housing at the city's edge, since they have the alternative of building on adjacent vacant land for which site rent is zero. At the city's center, development reaches a density of Q_0 apartments per building lot and site rent (the saucer-shaped area above the marginal cost curve) is at a maximum.

We have shown that the model in Figure 6.3 generates a skyline that rises higher and a bid-rent curve for housing sites that grows steeper as we approach the center of the city. Since the model assumes housing to be the only land use, the bid-rent curve for housing sites is by assumption also the site-rent gradient for the city. It displays the characteristic shape found in empirical studies of land value in major cities.

THE EFFECT OF COMPETING LAND USES ON URBAN FORM

The analysis developed above can now be modified to cope with something closer to the full complexity of land uses in a typical city. Although in principle the argument could be extended to any number of uses, let us for convenience combine city functions into three major groups and see how competition among them for sites will establish a land-use pattern and site-rent gradient for the city. The first group, which we will call "central office functions," includes such activities as corporate headquarters offices, banks and other financial institutions, and law and accounting firms, all of which are complementary in providing high level business services that require frequent daily contact between firms. The second group, which we will call "ancillary services," includes such categories as office equipment, parts and supply houses, printing shops, maintenance and repair firms, and telephone exchanges, whose function is to provide routine rather than high level services to the central office sector. The third category is housing.

We assume that each of the business sectors operates under conditions of perfect competition. We continue to posit a city on a flat transport surface so that movement is equally costly per mile in any direction. Within this city, central office functions and ancillary services, since they require frequent contact between firms, will exert a mutual attraction and will therefore locate close together rather than occupying scattered sites. It follows that the city will have a well-defined business district and that the center of this district will be the point offering greatest accessibility to other firms. The question to be answered is: How will the three kinds of land use locate in relation to this central point? The answer turns out to be that when various uses are competing for sites, the one that can pay the highest site rent at each particular location will come into possession there. Consequently the first point to investigate is: What determines the spatial pattern of demand for sites by each use?

To simplify the argument we make these additional assumptions: First, as in the housing case analyzed above, buildings are constructed and operated by developers who rent sites from landowners and in turn lease shelter space to tenants. Second, all structural types have similar cost characteristics. More specifically, all display identically increasing marginal and average cost per unit of floor space per year for building and operating structures of increasing height on lots of standard size. Thus the left-hand panel of Figure 6.4 shows cost curves resembling those of Figure 6.3. However, the horizontal scale to the left now measures quantity in square feet of floor space per structure rather than in number of apartments. As before, increased quantity (i.e., increased output per site) is obtained by building higher.

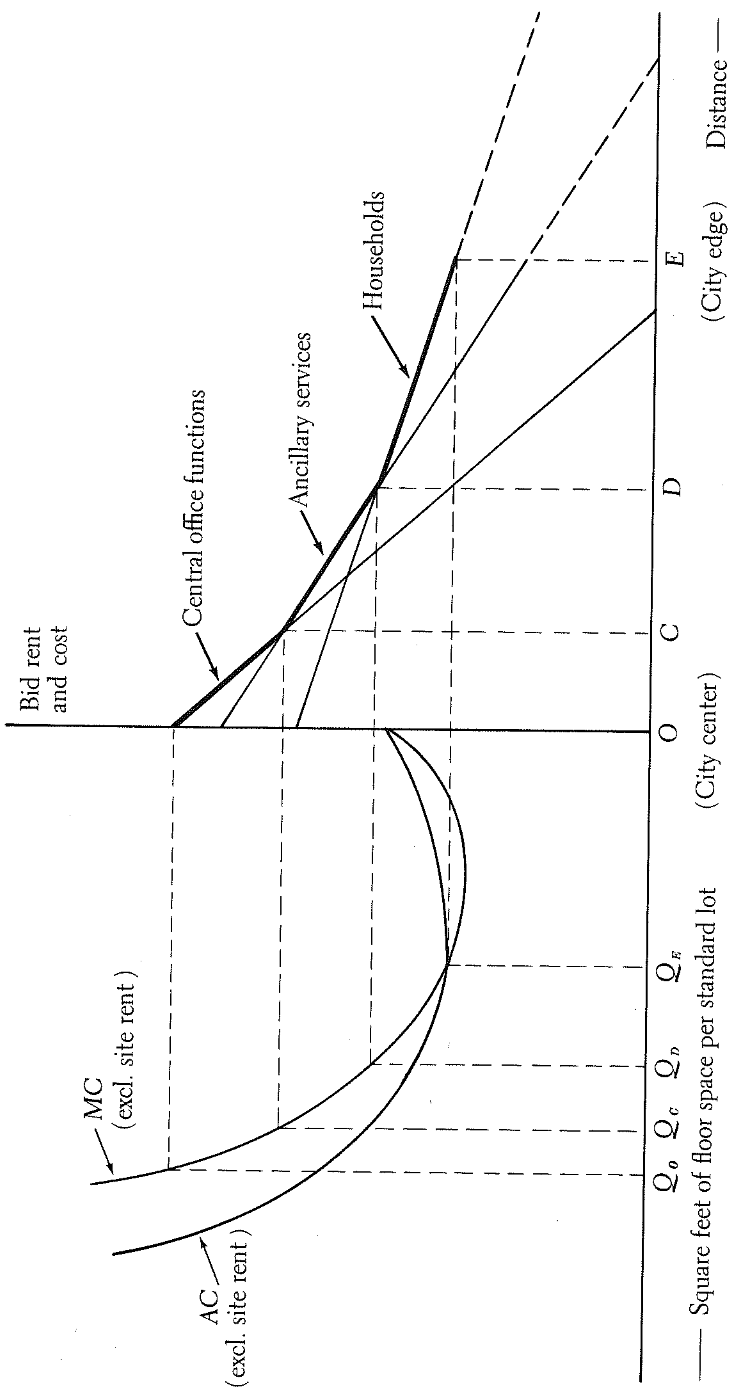
In the right-hand panel of Figure 6.3 we drew a bid-rent curve of households for apartments, showing the maximum price that would be paid at each location as a decreasing function of distance from the center. The curve represented utility-indifferent positions for individual households. In the same fashion, bid-rent curves can be drawn for floor space to be occupied by each of the other types of activity, and each of these will show the maximum rent that can be paid consistent with earning a normal competitive return. This rent will decrease with distance from the city center.

First consider central office functions. Each firm in this group is heavily dependent on daily face-to-face contact between its own executives and their counterparts in firms with which it deals. The closer each firm can come to location at the center of the business district, the less costly to it in time and transportation outlay will be the task of maintaining contact with other firms. Thus a central location will reduce the firm's own costs (other than rent). For the same reason it will reduce the cost to others of maintaining contact with that firm. Hence a central location will also increase the firm's sales volume. Thus the rent per unit of floor space that a central office firm can pay while still earning a normal return will increase with proximity to the center both because transportation and communication costs will be decreasing and because sales will be increasing as distance from the center diminishes.

Much the same argument applies to the group of ancillary services. Firms in this group are in business to supply services to the central office firms (and to one another). They, too, will find costs (except rent) decreasing and sales increasing as distance from the center diminishes. Thus the rent they can pay per unit of floor space while still earning normal returns will also rise with proximity to the center.

Would the bid-rent curve for floor space be steeper for central office

FIGURE 6.4



functions or for ancillary services? What determines its steepness in either case? Three factors can be distinguished. First, its steepness is greater, the greater the cost per unit of output of maintaining contact with the center. Second, its steepness is greater, the larger the number of units of output the firm produces per square foot of floor space occupied. Third, it is steeper, the less readily other inputs can be substituted for floor space in the production process as the price of floor space rises.[7]

It seems probable that central office functions would have a steeper bid-rent curve than ancillary services. First, since they are less routine and standardized than ancillary services, they are likely to require more frequent personal contact between firms. In addition, since contact by central office firms frequently involves highest echelon executives, while contact for ancillary service is more apt to be carried out by lower-salaried personnel such as salesmen, truck drivers, or repairmen, it is probable that travel time would be valued far more highly for the former group than for the latter. Second, the quantity of output per unit of floor space is likely to be larger for a central office function than for an ancillary service. Comparable units of "output" are difficult to conceive of in this instance, but the point is that central office functions are, in some sense, able to make more intensive use of space than are ancillary services.

We show the bid-rent curve for households as the least steep of the three. The household typically has less need for central location than do business firms. Whereas the firm wishes to facilitate numerous daily contacts with customers and suppliers, the household need make only one daily journey to the center for each working member. (We have not introduced shopping activity into the model.) To shorten these daily trips, households are willing to pay higher rent, as shown by a bid-rent curve that rises toward the center, but they are not willing to pay as much for centrality as do business firms. Thus the household curve lies below the inner (more central) portions of the other two.

The spatial distribution of the three land uses within the city is determined by the relationships among their bid-rent curves. At each location building owners, in order to maximize net income, rent space to the highest bidder. Hence the use with the highest bid-rent curve takes over occupancy at each point. In Figure 6.4, central office functions occupy the segment from O to C, ancillary services locate between C and D, and households occupy the segment from D to the edge of the city at E. Thus the highest portions of the three bid-rent curves (as indicated by the heavier line) become the rent gradient for the city. In this instance the gradient refers to rent per square foot of shelter space, but, as we have

---

[7] For a more extended discussion of these factors see Hugh O. Nourse, *Regional Economics* (New York: McGraw-Hill, 1968), pp. 96–110.

argued above, a unique land-rent gradient is, under the assumed conditions, associated with it. Both gradients become steeper as they approach the center. If we rotate either one about the city center we generate a corresponding rent surface. Looking down at such a surface from above we would see that the city is circular, that central office functions occupy the center and that the other land uses are ranged thereabout in concentric rings.

Interdependence Through Competition for Sites

The arrangement of land uses produced by the model depicted in Figure 6.4 forms an interdependent system. Not only are the several activities related to one another through the usual linkages of complementarity and substitution; they are related also by their competition for sites. In a more elaborate model that comes closer to articulating the general equilibrium nature of the problem, one would link the demand for sites by each activity to the quantity of output of each that could be sold in the markets served. In such a model the values of all supply and demand variables would have to be solved for simultaneously, since the site area occupied by each activity influences its output from the supply side, while at the same time the area that can be occupied by a given activity in competition with other uses depends on the demand for its output, the supply of complementary factors, and so on. The partial equilibrium model employed here can suggest the nature of these interdependencies, although it does not incorporate them explicitly.

For example, the outcome shown in Figure 6.4 was assumed to be an equilibrium division of land uses. In arriving at that division it was assumed implicitly that the quantity of central office functions that could be produced in the circular area of radius OC was just that quantity which could fetch a price sufficient to enable central office firms to exert the demand for space represented by their depicted bid-rent curve. Analogous assumptions held for the other land uses. Now suppose that an exogenous increase in the demand for central office services occurs. The price of such services would rise, and producers would seek to increase output. Their demand for sites—which is a "derived demand" based on the demand for their output—would increase. We could represent that increase by shifting their bid-rent curve up and to the right, which would cause the point of intersection of their bid-rent curve and that of ancillary services to move rightward. The immediate result would be an increase in the area occupied by central office functions at the expense of the ancillary service area.

But the chain of effects does not end there. If the area occupied by ancillary services shrank, output would fall and prices would rise. At higher prices for output, this industry, too, could bid more for sites. Hence the bid-rent curve of the ancillary service group would also move up and to the right (and thus would have an analogous effect on household demand for sites). The final equilibrium of the system would show that the margin of the central office area had moved out, but by less than would have occurred if the bid-rent curve of ancillary services had not thereby been forced up. Likewise the margin of the latter ring would have moved out, but by less than would have occurred had the household curve remained fixed. Finally, the upward shift of the household curve would have pushed the edge of the city farther from the center. All these effects are the result of direct competition between uses in the land market. We need not here trace out the effects of additional interdependencies, such as the increased demand for ancillary services when central office functions expand or the increased population that would be attracted into the household sector by these expanding industries.

We see, then, that a change in one part of the system causes a spreading wave of effects that in some degree alters the entire land-use pattern. Just as central place theory explains how cities are held in place, so to speak, by the location of all other cities, so, too, a fully developed model of intraurban land use would have to explain how the whole array of urban activities is mutually held in place by intraurban forces of attraction and repulsion operating through competitive land markets.

One might well ask whether our manner of combining the bid-rent curves in the model depicted in Figure 6.4 is not simply arbitrary. Why, for example, do the flatter bid-rent curves necessarily lie below the steeper ones at the center? The answer is that no other arrangement provides a stable equilibrium. If we were dealing with linear bid-rent curves for shelter space (i.e., each with constant slope throughout) and the curve that was highest at the center were also flattest, then it would be higher than all other curves throughout its length, and the city would be entirely devoted to that single land use. This is true a fortiori if the bid-rent curves are concave upward. But suppose one of them were concave downward? Then it might be higher but also flatter than the others at the center. This case can be ruled out, however, since it implies that the cost per mile of transportation to the center increases with the length of the journey, whereas the general tendency in transportation is for cost per mile to remain constant or else decrease with distance covered.

If these arguments seem to lean too heavily on mere geometry, consider the economic logic that lies behind them. The slope of a bid-rent curve measures the benefit that accessibility confers on a given activity: the greater their need for accessibility in terms of reduced cost or increased sales, the more firms of a particular type are willing to pay in higher rent to move one mile closer to the center. The more they are willing to pay per mile, the steeper their bid-rent curves. The slope of the bid-rent curve of a particular activity, therefore, measures the benefit that accessibility confers on it. Industries that can benefit the most from accessibility have the steepest curves and occupy the center. Hence the model behaves efficiently in the economic sense: it allocates the scarce resource of accessibility to those who can make the most productive use of it.

The Doctrine of "Highest and Best Use"

Consistent with the foregoing argument, it is often said that a competitive real estate market allocates urban sites to their "highest and best use." Competition puts sites in the hands of the highest bidder. The highest bidder is the one who can make the most economically productive use of the site. The market operates so as to maximize rent from each site. We have already shown that site rent equals the marginal product of land. Hence the market also maximizes the contribution (i.e., marginal product) that each site adds to total output. It is in that sense that "highest" is also "best" from the viewpoint of society as a whole.

As we will see in Chapter 13, however, the process by which one use *succeeds* another on a given site is far more complicated than the above passage suggests. If a cleared site is thrown on the market it will obviously be sold to the highest bidder, who will then construct a building on it that represents the highest and best use of the site. If, however, a site has an old building on it that is still capable of rendering service, that old use may be sufficiently profitable to persist on the site even though it is not the kind of building that anyone would now construct if the site were already cleared. Frequently observed examples of this sort are the old four-story commercial buildings that stand cheek-by-jowl with skyscrapers in the CBD or the small, walk-up apartment houses scattered among tall, modern elevator structures in a residential district. These cases are not exceptions to the doctrine of highest and best use. Properly interpreted, the doctrine comprehends them. However, the matter will not be taken up in detail until we analyze the economics of land-use succession in Chapter 13.

EXTERNALITIES AND LAND-USE ZONING

A major qualification to the argument that competitive land markets operate efficiently in the assignment of sites must now be introduced. We have treated land uses as though they were independent of one another except for those connections made through market transactions. Thus we have ignored external, or neighborhood, effects. These arise when activity at one site confers benefits or imposes costs on the occupant of another site for which no fee can be charged or no recompense collected. For example, a beautiful garden in front of one house produces a free aesthetic benefit for neighbors and passers-by, while the noise and fumes from a boiler factory impose unrequited damages on the occupants of nearby sites. Such effects are especially likely to occur in densely built-up urban areas.

If the owners of the boiler factory could somehow be made to bear its external costs, they would either contrive to reduce the output of noise and fumes or else move to a more remote location where such emissions would cause less offense. Land-use zoning arose as an attempt to meet this problem, not by inducing occupants of sites to internalize the costs they impose on others, but by direct regulation of land use. Although some forms of land-use regulation appeared earlier, comprehensive zoning in the United States is usually dated from the adoption of a zoning resolution by New York City in 1916.[8] Zoning was intended to minimize what we now call "externalities" in two ways. First, incompatible uses were to be kept from impinging on one another. Second, regulation was to prevent overintensive development of one site from imposing burdens on its neighbors. Incompatible uses could be kept apart by zoning certain areas for residential development to the exclusion of all industry and others for residential, commercial, and light industrial uses, while confining truly "noxious" activities, such as boiler factories or stockyards to peripheral locations. The same zoning ordinance could prevent individual improvements from blocking the light and air of their neighbors or imposing other burdens on them by regulating the height and bulk of buildings or requiring open space along lot boundaries.

Land-use zoning was quickly accepted in the United States after the Supreme Court in 1926 upheld its constitutionality in the *Euclid* case. Zoning ordinances are adopted and administered by local government, under authority granted to them by state law. The preparation and periodic review of such ordinances is often one of the principal functions of a city planning commission.[9] Exceptions (known as "variances") to the specific requirements of an ordinance are usually provided for preexisting uses that do not conform, and additional variances can generally be granted by a board of standards and appeals.

In recent years, many economists have begun to question whether zoning, as it is practiced in the United States, is either an efficient or an equitable way of regulating land use. Examining the issue of efficiency,

---

[8] See John Delafons, *Land Use Controls in the United States* (Cambridge, Mass.: M.I.T. Press, 1969), pp. 16–31.

[9] For a description of zoning from the viewpoint of the city planner, see Anthony J. Catanese and James C. Snyder, eds., *Introduction to Urban Planning* (New York: McGraw-Hill, 1979), ch. 10; and Richard F. Babcock, "Zoning," in Frank S. So, et al., eds., *The Practice of Local Government Planning* (Washington, D.C.: International City Managers' Association, 1979), pp. 416–43.

Mills, for example, points out that land-use zoning purports to control nuisances such as noise pollution when, in fact, it merely moves them about. While not opposing some degree of use separation through zoning, Mills argues that antipollution policies aimed directly at externalities would often be more effective and would render some traditional zoning controls redundant. Appropriate policies would include greater reliance on such devices as effluent fees or other charges that encourage economically efficient marginal adjustments in the behavior of firms and individuals, in place of the all-or-nothing system of toleration or prohibition through zoning.[10]

The all-or-nothing structure of zoning not only prevents efficient marginal adjustments but also contributes to the politicization of the zoning process. Whenever a particular use for which there is strong market demand has been forbidden by zoning—say the construction of apartment houses in a neighborhood zoned exclusively for single family homes—a powerful incentive is created for an aggressive developer to profit by obtaining a zoning variance that breaks the restriction. Political forces are brought into play. Zoning then becomes a game in which the prize is the potential private gain from changing the initial rules.

The most frequently voiced objection to zoning on grounds of equity, however, is not that the zoning process itself is highly political, but rather that it can be used by one class of citizens to the disadvantage of another. For example, by manipulating zoning regulations, well-to-do suburban towns can effectively prevent families of low or moderate income from moving in. This sort of "exclusionary zoning" will be examined in detail in connection with housing policy in Chapter 13.

INTRODUCING THE EFFECTS OF CHANGE AND GROWTH

The land-use model depicted in Figure 6.4 is wholly static. Ignoring time and change, it shows us what the equilibrium pattern would be if a city were suddenly to be built *de novo* under given conditions. It does not tell us how land-use patterns evolve as cities age or as the things assumed constant in the model—especially population, income level, and technology—change through time. It leaves out of account all the dynamic forces

---

[10] Edwin S. Mills, "Economic Analysis of Urban Land-Use Controls" in Peter Mieszkowski and Mahlon Straszheim, eds., *Current Issues in Urban Economics* (Baltimore: Johns Hopkins University Press, 1979), pp. 511–41. The article includes a useful bibliography. For a more extended economic analysis see David E. Ervin, et al., *Land Use Control: Evaluating Economic and Political Effects* (Cambridge, Mass.: Ballinger, 1977).

of urban evolution that for better or worse prevent the achievement of any final equilibrium. Because structures and—even more so—the underlying framework of streets and utilities are long lived, the pattern of land uses that exists at any actual time is never the same as the optimal pattern that could be produced by a wholly fresh start at that moment. The aging of structures and the process of land-use succession on given sites will be examined in detail in Chapter 13. At this point we wish to analyze the general, or macrolocational, effects on land use of changes in transportation cost, population, and income.

The consequences of such changes have been worked out very clearly by Alonso, and the following discussion is based largely on his work.[11] Although Alonso's study covers business and agricultural as well as residential patterns, he simplifies the formal analysis of the effect of changes in technology and the like by restricting it to the latter sector. (The results are fundamentally the same for urban business uses as well.) Alonso's model, like the highly simplified one used above, assumes a transport surface and a mononuclear city. It is far more complex, however. Among other things, it explicitly rejects the assumption of constant residential lot size. Rather, lot size is one of the variables to be solved for. What follows below is not the Alonso model itself but one of its applications.

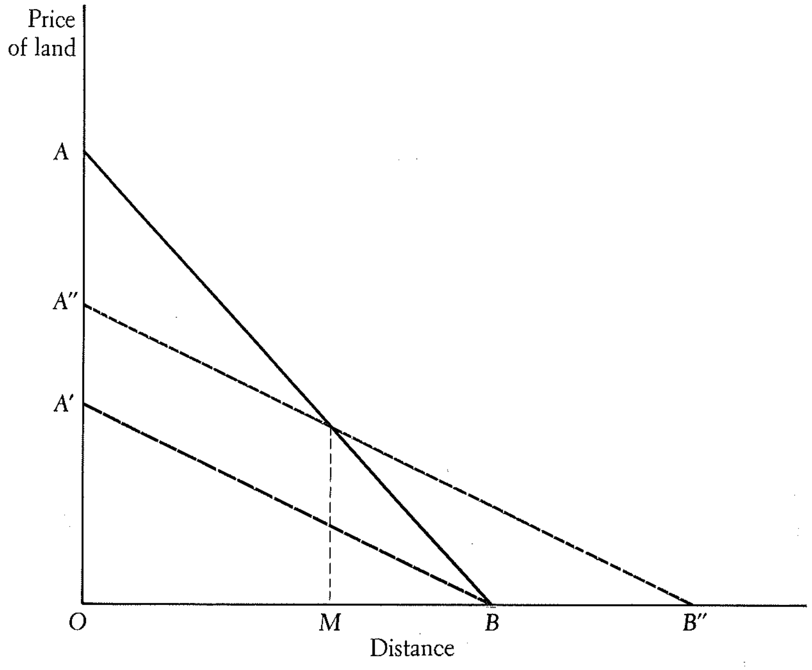
The Effect of an Improvement in Transportation Technology

Figure 6.5 shows the bid-rent curve of households for land (or, as Alonso calls it, the price structure for residential land) in a metropolis with its business and employment center at O. The land units in terms of which price is expressed may be square feet, acres, or what have you. Initially, the price structure is given by line *AB*. *OA* equals rent at the center, and point *B* marks the edge of urban settlement. Now suppose that an improvement in technology takes place that reduces the time and/or money cost of transportation from the outlying areas to the center. Rent at the center is based on the saving in transportation cost obtained by locating there instead of at the city's edge. Accordingly, rent at the center will be reduced by the technological improvement, other things remaining the same. More specifically, if lot size were held constant the outer edge of settlement would remain at *B* and the bid-rent curve would fall to *A′B*. But with the price of land reduced, lot size will *not* remain constant. Alonso's model incorporates the important fact that to the householder space is a consumer good as well as an impediment to access. Other things being equal, if the price of land drops, householders will increase their consumption of it: lot size will increase. That, in turn, means that the area needed to

---

[11] Alonso, pp. 105–16.

FIGURE 6.5



Source: Adapted from William Alonso, *Location and Land Use* (Cambridge, Mass.: Harvard University Press, 1964), fig. 32, p. 112.

house a given population will also increase. The margin of settlement will move out—say, to *B″*. Instead of falling to *A′B*, the bid-rent curve will shift to a position such as *A″B″*.

We know that *A″* must be above *A′*—in other words, that rent at the center falls less than it would have if lot size had remained constant—because when the city's edge moves from *B* out to *B″*, the transportation cost saved at a central location increases, and therefore so does rent at the center. It is interesting to note that although the new bid-rent curve *A″B″* shows rents lower than before at the center, it also shows them to have increased beyond some point, *M*. Why should households beyond *M* be paying higher rents after the improvement in transportation than they paid before? The answer is that people living farther out than *M* could pay more rent than before and still be better off than they were because the technological change has allowed them to gain utility through reduction in the time and/or money cost of transportation.

The Effect of an Increase in Population

The effects of population change also are easily deduced from Alonso's model. An increase in population will increase the demand for residential sites. In order to accommodate the new households, the margin of the city will move outward. As it does so, the cost of transportation from the edge to the center will increase, and the bid-rent curve will shift upward and to the right. For example, an increase in population, while transportation technology and income remain unchanged, will cause the bid-rent curve to rise from a position such as *A′B* in Figure 6.5 to something like *A″B″*. The result will be higher land costs for all households and a tendency to reduce lot size. Hence the city's area will increase less than in proportion to the increase in population. Thus, as Alonso points out, a rise in population, other things being equal, will cause an increase in both land prices and density of settlement as well as in the physical extent of the urban area.

Figure 6.5 also enables us to compare expected land costs in cities that differ in population size. The diagram predicts that land prices, and therefore housing costs, will generally be higher in large than in small cities, and that is, indeed, the case. As we pointed out in Chapter 3, these higher costs, in turn, help to explain why the cost of living rises with city size. (See the section on "Income, Well-being and City Size.")

The Effect of an Increase in Income

We next examine the impact of changes in average household income. Alonso deals with the effect of differences in income among households rather than with the effect of a change in its average level. Once the former effect is established, however, we can easily transform the argument to describe the latter.

The facts are not in doubt. In U.S. cities wealthier families tend to settle near the periphery, while poorer households remain close to the center. Thus the gradient of income rises with distance from the center. This is crudely verified, later, in Tables 10.5 and 11.7. Within SMSA's median family income is lower in the central city than in the suburbs. (It is equally clear that the U.S. pattern is not a universal norm. In many cities of Latin America, for example, the well-to-do live near the center, while the poor occupy peripheral sites.[12])

As we shall see, there are two alternative explanations of the U.S. pattern. Both are plausible, but they have very different implications for

---

[12] See Leo F. Schnore, "On the Spatial Structure of Cities in the Two Americas," in Philip M. Hauser and Leo F. Schnore, eds., *The Study of Urbanization* (New York: John Wiley, 1965), pp. 347–98.

housing and renewal policy in the central city. In what follows, both are examined in detail.

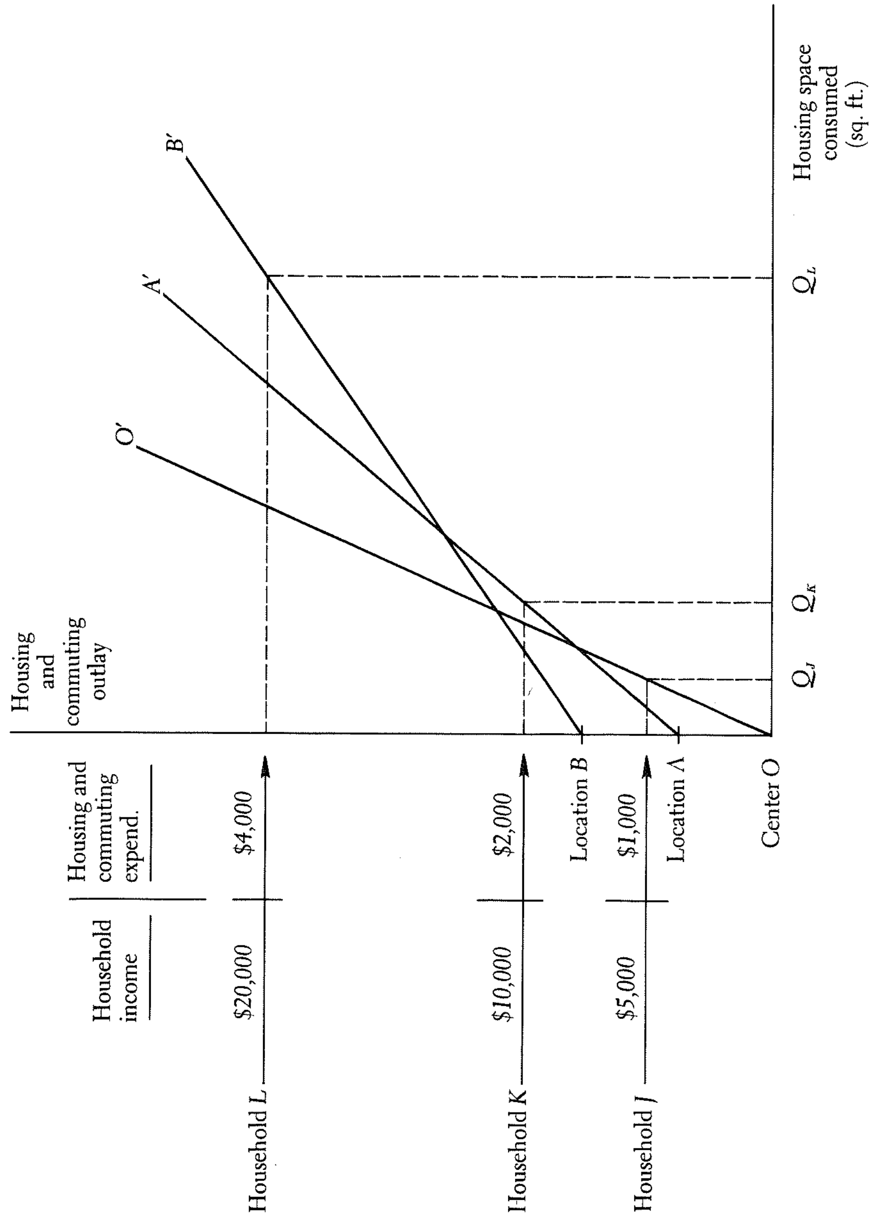
Alonso argues that higher income, in and of itself, can lead families to choose suburban rather than central city locations. By close theoretical reasoning he shows that wealthy families will tend to have bid-rent curves less steep than those of poor families and therefore settle on large lots of relatively cheap land toward the city's edge, while the poor tend to occupy very small portions of higher priced land near the center. The argument is complex. An important factor in it, however, is the cost of commuting. Assume to begin with that the cost of commuting increases directly with distance but does not vary with income. (We will reconsider this assumption in a moment.) On the other hand, the cost of housing per square foot declines with distance because the land component becomes cheaper with movement away from the center. Households therefore face the following choice: at locations close to the center commuting costs are low but the unit costs of land and housing are high; at locations farther out commuting costs are high, but the unit costs of land and housing are low. For the poor family, the increase in commuting costs as distance increases will diminish rapidly the small fund of income available for housing. Consequently the poor cannot bid much for locations where commuting is expensive. On the other hand, since commuting costs are invariant with income or quantity of land occupied, the rich, who desire ample housing space and are prepared to spend large sums on housing, find the barrier of commuting costs rather inconsequential and can bid higher prices than the poor for land at distant locations. In choosing such locations the rich gain more by consuming larger quantities of cheaper housing than they lose by paying additional transportation costs. (Of course, Alonso recognizes that individual tastes are not uniform. Some wealthy families with a strong aversion to commuting and a weak preference for added space will always continue to live in luxury housing near the center.)

Figure 6.6 may help the reader to visualize the outcome just described. Consistent with the above explanation and with the earlier treatment of housing and transportation outlays in Figure 6.3, it illustrates the way in which income influences household locational choices.[13] In this interpretation households regard commuting outlays as part of the cost of occupying housing. At all levels of income they are assumed to budget 20 percent of income for housing and commuting costs combined. This implies that the income elasticity of demand for housing is 1.0, since housing expenditures always rise in proportion with rising income. Each

---

[13] Figure 6.6 is an interpretation of the income-distance effect that does not appear in Alonso.

FIGURE 6.6



household seeks the location at which it can obtain the largest quantity of housing space for the outlay budgeted. Thus households do not have preferences among locations as such but choose solely on the basis of housing cost per unit of space.

In Figure 6.6 quantity of housing consumed per household is measured in square feet along the horizontal axis. The vertical axis measures annual dollar outlays for housing and commuting. Cost curves relating outlay to square feet consumed are shown for three locations. The commuting cost at each location is given by the intercept of the cost curve on the vertical axis. Thus annual commuting cost to the center of the city from location *A* is *OA* and the curve *AA′* relates housing outlay to square feet occupied at *A*. Likewise, *OB* is the cost of commuting at location *B*, which is farther out than *A*, and the slope of *BB′* measures the cost of housing at *B*. The slope of *OO′* indicates housing cost at the center, where commuting costs are zero. Since commuting costs are directly proportional to distance, intercepts on the vertical axis can be read to indicate relative distance of each location from the center. As indicated by the diminished slopes of successive curves, cost per unit of housing space declines as distance increases because land costs fall with distance.

Figure 6.6 illustrates locational choices for households at three income levels. The outcome for each household is read on the diagram by following that household's horizontal income line to the right and finding its intersection with the housing cost curve that lies farthest out along the quantity scale. Household *J* with an income of $5,000 budgets $1,000 a year for housing and commuting. At that level of spending the largest number of square feet is obtained by living at the center of the city, where commuting costs are zero. Household *K*, with a budget of $2,000 for housing and commuting, finds that location *A* offers the most space, while household *L*, spending $4,000 per year, settles still farther out at *B*. If housing cost curves for all locations within commuting range of the center were plotted, the successive portions that lie farthest to the right would generate something approaching a smooth curve from which one could read the location that would be optimal for households at every income level. Under the conditions assumed here, household choices would yield an income gradient rising continuously with distance from the center. The forces producing that result can be observed in Figure 6.6. Housing outlays double as we move up the income scale from *J* to *K* and from *K* to *L*. Likewise distance doubles from *A* to *B*. But for each doubling of income or distance, the market offers more than twice the quantity of housing space: Q_K is more than double Q_J, and Q_L more than double Q_K.

This interpretation of Alonso's argument can readily be extended to cover the effect of an increase in the average level of income as well as the effect of income differences among households. As the income of American households increases over time, they move gradually up the vertical scale of Figure 6.6. Consequently, the proportion of the population that chooses to take up suburban living increases. This means that even if population, tastes, and the technology of transportation were constant, rising living standards alone would suffice to increase the suburban population and diminish that of the inner cities. This process would also produce the trend in geographic income differentials that we do in fact observe and will discuss at length in later chapters: average family income is rising far faster in the suburbs than in the central cities.

Income and Travel Cost

The results obtained from Figure 6.6 are unambiguous because it was assumed explicitly that commuting cost is a function of distance but not of income. This was a convenient simplifying assumption that must now be given up. As already explained in Chapter 5, travel cost in general consists of two parts: first, the money cost of making a trip; second, the nonmonetary cost comprising the value of time lost and of discomfort undergone. It is reasonable to assume that the money cost is proportional to distance—that is, so many cents per mile—and that it does not vary with income. The time cost of travel, however, is a different matter. As we shall explain in Chapter 8, empirical studies indicate that individuals value travel time at some fraction of their wage rate. The perceived time cost of covering a given distance will therefore rise with income, although not necessarily in proportion to it.

In models of the Alonso type the effect of income on the locational choices of households depends upon whether the demand for space or the marginal cost of travel increases faster with income. If the income elasticity of demand for space is greater than the income elasticity of the marginal cost of travel, then the rich will tend to locate farther out than the poor. This is the case depicted in Figure 6.6, where the former elasticity is assumed to be fairly high, while the latter is assumed to be zero. If, however, the latter elasticity were to exceed the former, the burden of distance would increase faster with income than would its advantages. The income-distance pattern would then be reversed, with the rich living at the center and the poor toward the edge.

Alonso, Muth, and others, who argue that higher income per se explains the peripheral location of the well-to-do in U.S. cities, do so from a belief that in the United States the income elasticity of demand for space exceeds the income elasticity of the marginal cost of travel. This belief gains plausibility from the fact that the total cost of travel for any individual will always be less responsive to income than the time cost, because the total includes a money cost component that depends only on distance.[14]

Recent work by Wheaton, however, questions whether what he calls the Alonso-Muth spatial income effect is really an important determinant of residential location. Using data for a sample of households in the San Francisco Bay area, he estimated the effect of income on locational choice. He concluded that it was relatively weak because the estimated income elasticities of land consumption and of travel cost proved to be very similar in magnitude.[15] Further empirical work will be required, however, before this issue can be resolved.

If the income effect per se is as weak as Wheaton's findings suggest, then the suburbanization of middle and upper income families in the United States must, of course, have had other causes. Which causes, in fact, produced the present pattern is an important question, since the answer has implications for policies intended to stimulate central city revival.

An Alternative Explanation of the Spatial Income Pattern

The principal alternative to the Alonso-Muth explanation is the view that the outward movement of the middle and upper classes is the result of the growth, development, and aging of the central city over time. Based on observations in Chicago, this theory was first put forward by the sociologist Ernest W. Burgess in the 1920s and has been highly influential.[16] It holds that the wealthier classes, who originally lived in fashionable districts near the city center, moved outward for two reasons. First, the expansion of the CBD as the city grew encroached on their close-in residential zone, destroying its amenities. Second, since the close-in housing was the oldest, it was also the first to become obsolete. Because it is easier to build new on vacant land than in built-up districts, the wealthy tended to move to the periphery to construct another round of up-to-date housing. Older housing was abandoned to poorer classes, and the oldest of all (adjacent to the CBD) to the immigrants whose arrival fed the city's growth.

This theory of city form has come to be known as the "concentric zone theory," because Burgess described four distinct types of housing, which he believed were grouped in concentric rings around the CBD. He identified the zones and their residents as follows (moving from the center outward):

1. The Central Business District (in Chicago, "The Loop")
2. The Zone in Transition: as the CBD grows, business and light manufacturing encroach on old slums and rooming houses, making this the least desirable residential area
3. The Zone of Workingmen's Homes: inhabited by laborers who have escaped from the deterioration of Zone 2 but wish to live close to their central workplaces
4. The Residential Zone: a restricted or exclusive district consisting of high-class apartments and single family homes for the middle and upper classes
5. The Commuters' Zone: suburban areas and satellite cities, outside the city limits but within commuting distance of the CBD

Burgess's view that the spatial arrangement of urban social classes is explained by the growth and aging of the city can easily be brought up to date. European immigrants were replaced by poor ethnic minorities in the deteriorating housing at the city's core. Racial tensions in the inner city grew, and poverty brought crime and fiscal distress, further encouraging the exodus of the white middle class, which sought relief from such problems in suburban communities segregated by income class as well as by race. In Wheaton's concise summary this view holds that "the suburbanization of America's middle and upper classes is a response to housing market externalities and the fiscal incentives of municipal fragmentation."[17] (These fiscal incentives will be carefully examined in Chapter 15.)

Policy Implications of Alternative Theories

Alternative explanations of the suburbanization of the middle and upper classes clearly have different policy implications. For example, if housing obsolescence at the core is an original cause of the exodus, then policies that stimulate inner city housing renewal might stem and eventually reverse the outflow and thus encourage general central city revival. This argument was used to support federal subsidies for urban renewal projects, including the construction of new luxury housing at the center, in the 1950s and 1960s. (See Chapter 13.) The case is entirely different, however, if the preference of households for space-consuming

---

[14] Muth, p. 31.

[15] William C. Wheaton, "Income and Urban Residence: An Analysis of Consumer Demand for Location," *American Economic Review*, September 1977, pp. 620–31.

[16] Ernest W. Burgess, "The Growth of the City," in Robert E. Park, Ernest W. Burgess, and Roderick D. McKenzie, *The City* (Chicago: University of Chicago Press, 1925).

[17] Wheaton, p. 631.

suburban living increases directly with income. In that event, the outward movement of the wealthy is simply the result of their rising living standards, and urban renewal in the older central cities will be relatively ineffective in stemming their outward migration or attracting a return flow.

Much the same reasoning applies to a diverse group of other policies that attempt to overcome structural obsolescence or race and class antagonism in the central city. These policies may be desirable for other reasons, but they are less likely to have the additional effect of attracting the middle and upper classes back to the city if the spatial income effect described by Alonso and Muth has, indeed, been a major cause of their suburbanization.

Recapitulation: The "Automobile Effect," the "Overflow Effect," and the "Income Effect"

It is worth noting that with regard to the causes of dispersion within metropolitan areas, the inferences drawn from Alonso's formal land-rent model coincide with those cited in the essentially descriptive-historical analysis of the same process we presented in Chapter 3. Referring back to Figure 6.5, assume for convenience that *M* marks the fixed boundary between a central city and its suburbs. Formal theoretical analysis then supports the following conclusions:

1. An improvement in transportation, other things remaining the same, allows the margin of urban settlement to move out. If total population remains constant this causes population to rise in the suburbs and fall in the central city. This is the "automobile effect" of Chapter 3.
2. An increase in total metropolitan population, other things remaining the same, also pushes the margin of settlement out. It leads, by way of reduced lot size, to greater density and therefore to a larger population in the central city as well as in the suburbs. However, since the suburbs start with a smaller population base, the percentage increase recorded there will be greater than in the central city. This is the essence of the "overflow effect" cited in Chapter 3.
3. An increase in the average standard of living, other things constant, reduces the demand for central sites and increases the demand for more distant ones. Average lot size increases, and again the margin of settlement moves out. With no change in total numbers, this means a smaller population in the central city and a larger one in the suburbs. This is the "income effect" of Chapter 3.

We see then that each of the three effects, taken separately, leads to increased suburban population, while two reduce and one—the population effect—increases central city numbers. Since early in the twentieth century all three have operated simultaneously. What were the consequences for the pattern of settlement? As we showed in Table 3.2, the growth rate of central city population was at its height in the first decade of the century. It slowed down more or less steadily thereafter and turned negative for some of the older central cities in the 1950s and 1960s. It thus appears that for a long time the strength of the population effect was sufficient to maintain some central city growth despite the strong impetus to dispersal from improved transportation and rising living standards. As metropolitan population growth slowed down, however, it was less and less able to offset the other two effects, which continued to make for dispersion throughout the 1970s. Consequently, central cities as a whole began to lose population after 1970.

The Effect of Higher Fuel Costs

Will the rising cost of energy now slow, or reverse, the long established trend of decentralization? In principle, higher fuel costs leading to increased costs of transportation would be expected to have the opposite effect to that illustrated in Figure 6.5 for an improvement in transportation technology. In a mononuclear city, site rent at the center is based on the saving in transportation cost gained by living there instead of at the edge. Accordingly, if transportation costs rise, rent at the center will increase and the rent gradient will become steeper, producing proportionate rent increases at all locations. But this will not be an equilibrium configuration. At higher land prices households will choose to consume less land. The edge of the city will move in. With density increasing, the population of the central city will rise. Assuming total population to be constant, the population of the suburbs therefore must fall. Thus in the model depicted in Figure 6.5, a rise in transportation costs, all other things remaining the same, would be expected to cause increased centralization within metropolitan areas.

However, it is necessary to add at once that reality differs significantly from the world depicted in Figure 6.5. Consequently, as we shall see in the next section, if centralization resumes as a result of higher fuel costs, it is likely to take quite a different form than would be suggested by that diagram.

TOWARD GREATER REALISM: TRANSPORTATION CORRIDORS, INTERPENETRATION OF USES, AND THE DECLINE OF MONONUCLEARITY

The model illustrated in Figure 6.5 is unrealistic in a number of respects. We have employed it because its basic implications are valid and important, but we must now recognize some real-world complications.

It must be admitted, first of all, that cities are not built on featureless transport surfaces. Topographical irregularities abound. Differences in elevation, view, wind direction, or proximity to natural features such as lakes, beaches, or mountains have an effect on locational choice. Equally important, transportation itself is not ubiquitous but is always channeled into corridors. These corridors themselves are often distorted by topographical irregularities, and the modes of transport vary according to corridor and purpose. Since major corridors offer superior accessibility along their own length, they generally command higher rents than minor streets or roads. Corridors, in effect, become centers.

Second, land uses are not arranged by the market into mutually exclusive districts, whether of concentric rings or any other shape. Not only are types of business intermixed within the CBD, but business and residential uses interpenetrate in both the central city and the suburbs. Retailing, for example, is found in all parts of the metropolis. This, however, is only the most obvious exception; other types of business, too, survive and prosper at a wide variety of locations.

Third, and most important of all, the dispersion of business activity has made metropolitan areas increasingly multinuclear. As we stressed in Chapter 3, urban evolution, driven by the force of changes in the technology of transportation, has moved steadily away from the nineteenth-century pattern of business activity concentrated at the center toward a much more decentralized arrangement. It was shown in this chapter that an improvement in transportation causes the city's residential zone to spread out, its population dispersing into a pattern of lower density settlement. The results are similar for business firms. Cheaper transport allows them to both obtain from suppliers and offer to customers as much accessibility as before in terms of time and money cost at a greater distance out from the center, where they can afford larger sites. Some firms, at least, will be induced by this change to move outward. The greater the possibility of substituting space for other inputs in the production process, the more likely they are to move. Thus automotive assembly plants or wholesale food distributors are more likely to disperse than are corporation law offices or advertising firms.

It should be emphasized that the business dispersion described here is not the spreading out of a discrete central business district at the heart of the city. Rather it is the dispersion of activity from that center toward the periphery; in other words, the suburbanization of industry. In the process new subcenters are created in the ring area. (To be sure, "satellite cities" already existed there.) These new centers often arise at intersections on the transport grid. Each exerts its own attraction for certain functions and emerges as a lesser peak on the urban rent surface.

Table 3.4 showed the extent of decentralization in twelve major SMSA's after 1948. The analysis in Chapter 3 and the population data in Table 3.4 also made it clear that the outward movement of jobs encourages, and in turn is encouraged by, the simultaneous outward movement of population. Since jobs are moving outward simultaneously with people, it follows that the suburbanization of population does not imply an increase in commuting. (See Chapter 3, section on "Changes in Work Trip Patterns.")

If renewed centralization now takes place under the pressure of higher fuel prices and rising transportation costs, it will occur in the context of metropolitan areas that have become distinctly multinuclear. In that setting, increased centralization is more likely to take the form of increased density in and around the new subcenters of the suburban ring than to appear as a substantial movement back into the old inner city. (For a more detailed analysis, see the concluding section of Chapter 9.)

EMPIRICAL EVIDENCE OF LAND-VALUE CHANGES OVER TIME

Empirical studies of changes in urban land-value gradients over time show that they have changed in the way our model would suggest under the impact of urban growth and decentralization. Mills estimated land-value gradients for Chicago from data gathered by Homer Hoyt for 1836, 1857, 1873, 1910, and 1928.[18] Mills' technique was to regress land value on distance from the CBD. He tested three forms of relationship at each date: linear in both variables, which fit poorly at all dates; log of land value against arithmetic distance—the equation for a gradient of the negative exponential type; and log of land value against log of distance. Since the negative exponential form fitted as well as any, we use it as the basis for this discussion.

Mills found that between 1836 and 1857 the slope of the land-value gradient increased moderately. The coefficient of determination, R^2, of the line of regression also increased, indicating that the negative exponential curve provided an increasingly good fit to the data. From 1857 to 1928, however, both the slope of the gradient and the value of R^2 gradually diminished. The flattening of the gradient is just what theory would have led us to expect as a result of successive improvements in the technology of intrametropolitan transportation. Moreover, Berry and Horton, building on Mills' analysis, have shown that the decline in the slope of Chicago's land-value gradient over that period is paralleled by a decline in

---

[18] Mills, "The Value of Urban Land."

the slope of the population-density gradient—again, just the combination that theory would predict.[19]

Mills' analysis also shows that while the slope of the land-value gradient was diminishing, its height at the center was steadily rising, so the whole gradient at each later date lay above its position at each earlier time. This rise in the level of the gradient is precisely what the mononuclear model would predict as a result of the vast increase in population during the period.

The decline in the value of the coefficient of determination, R^2, for the fitted gradients between 1857 and 1928 suggests, as Mills himself points out, that the mononuclear model has become less valid with the passage of time. As a result of the dispersion of business activity and the growth of other centers, distance from the CBD is gradually losing its once commanding power to explain intrametropolitan variation in site value. Economists are now developing more sophisticated approaches to the problem of intrametropolitan location, primarily through the use of computerized urban simulation models. The need for these complex techniques, however, does not render the simpler mononuclear theories useless as explanatory devices. Even in more advanced models the fundamental logic of business location decisions depicted in this chapter remains essentially intact: firms locate where they expect to maximize profits and are strongly influenced in so doing by the trade-off between accessibility to customers and suppliers on the one hand and the site rent they must pay to obtain that accessibility on the other. Despite the complexities of real-world location patterns, each class of urban activity does display a characteristic, measurable locational tendency. The observed complexity, in fact, is simply evidence that for most functions there remains a sufficiently wide range of feasible locations at going rent levels to produce considerable overlapping in spatial distribution.

THE ROLE OF SITE RENT UNDER DYNAMIC CONDITIONS

Whatever their differences in either structure or complexity, theories of intraurban location do not disagree on the crucial role of site rent as the free market's allocator of land among alternative uses. We opened this chapter with the analysis of rent as an allocator under static conditions. It is appropriate now to examine the role of rent under the impact of change and development.

---

[19] Berry and Horton, p. 302 and fig. 9.18. Chapter 9 of their volume reviews the extensive empirical work that has been done on population-density gradients and their change over time.

Consider the situation in a city with a rapidly growing population. Land value per acre in such a city will generally be rising, just as theory suggests. The rise in value and the growth of population stem from a single source: the economic advantages that the city offers in terms of individual income and business profit as compared with other places. To the extent that a given site confers valuable access to these gains, its owner can appropriate a share of the city's net advantages in the form of urban site rent. These net advantages, as we explained in Chapter 2, consist of the difference between the economies and the *dis*economies of agglomeration. Assuming perfect competition in factor markets and full geographic mobility of all nonland factors of production, landowners as a group would, in the long run, be able to appropriate the entire value of the city's net locational advantages. This would equal the difference between the aggregate returns that the nonland factors of production can earn when used in the city (net of any payments to compensate them for bearing the diseconomies of agglomeration) and the returns they could obtain if employed at a place where urban site rent was zero. The aggregate of site rent in a city is thus one measure of the net economies of location and agglomeration that the city offers.[20]

To the individual business that must rent space, high site costs per se are, of course, a *dis*advantage. As the city grows, some firms that found it profitable to locate there when rents were low may find it unprofitable to remain as they rise and will choose to move away. But the fact that site rents are rising (by which we mean rising relative to the general level of prices) in itself is sufficient evidence that the city's net advantages to land users as a whole have increased. Were it otherwise, bidders would not be pushing site rents up. Thus, from the point of view of society as a whole, high site rents should not be regarded as one of the "diseconomies of agglomeration." On the contrary, they measure the extent of the positive economies of agglomeration and location to be found there.

In practice the spatial redistribution of activity that occurs under the pressure of rising site rents may take forms other than the movement of whole firms. Haig long ago pointed out that business firms actually comprise many distinct functions that need not all be carried on at the same place.[21] As site rents in the CBD rise, firms often separate out those functions that do not require centrality—manufacturing, warehousing, and

---

[20] For an empirical study of the relation between aggregate land value (assumed to measure net economies of agglomeration) and city population size, see Matthew Edel, "Land Values and the Costs of Urban Congestion: Measurement and Distribution," *Social Science Information*, vol. 10, no. 6, pp. 7–36.

[21] R. M. Haig, "The Assignment of Activities to Areas in Urban Regions," *Quarterly Journal of Economics*, May 1926, pp. 402–34.

record-keeping, for example—and move them to less costly areas on the periphery or even to points outside the metropolis. In the end perhaps only the head office will remain in the CBD. A prominent recent example of this has been the tendency of book publishers, who are still heavily concentrated in New York, to move their distribution, storage, and billing operations to localities on the outskirts of the metropolitan area, where rent is lower, while retaining head office, editorial, and sales functions in New York City.

TRANSPORTATION PLANNING AND LAND USE

The argument of this chapter has focused exclusively on the way the private land market determines patterns of urban land use. No mention has been made of public planning or intervention, apart from zoning. This was deliberate but also unrealistic, for the land-use pattern in any city results from a combination of private and public decisions. In choosing where to locate, households and firms (the private sector) have to trade off the value of greater accessibility against the cost of obtaining it. In the real world, however, as opposed to the world of the "transport surface" assumed at the beginning of this chapter, the accessibility of any point is determined not only by its distance from the city center but also by its relationship to an existing transport network, and transport networks, including streets, highways, and the mass transit system (if any), are planned and paid for by the public sector. The actual urban land-use pattern thus results from interaction between the private sector, in which individual location decisions are made, and the public sector, which plans and develops a transport network that inevitably influences subsequent private decisions. In this chapter we have dealt only with the private side. In Chapters 8 and 9, which deal specifically with the economics of transportation, we will tell the other half of the story when we examine the impact of public policy on transportation and, through transportation, on urban form.