CHAPTER II

CITY FORMATION AND CITY SIZES

Complexity generally increases more rapidly than realism in model building. Although we understand many of the principles governing city formation, we do not yet have a model which includes any large part of our knowledge of real cities and remains simple enough to work with. We can, however, extend the basic model of Chapter I in several ways, and obtain interesting results.

In a sense, the open and closed cities of Chapter I hang in mid-air. We assumed either a given population or a given level of utility, without considering how that level came about. If we are interested in how these variables are determined, we need a general equilibrium model of an economy containing cities, not just a model of a single city. In this chapter we explicitly introduce a rural sector spread over on a featureless plain. The rural sector produces an agricultural good which is consumed by households in both the rural and urban sectors. Circular cities producing an urban good are sprinkled about on the plain.

In Chapter I we assumed that commuting is costly. The commuting costs and, in fact, transportation costs in general work against city formation. Concentration of production, for example, requires the transportation of products, workers, and material inputs. To obtain cities in our model, therefore, we must assume that the concentration of economic activities results in a technological advantage which at least exceeds the transportation costs incurred. Otherwise, production will take place where there are consumers, and the consumers, who find no advantage in working together, will spread out evenly to take advantage of all the available land.

Cities will arise in our model if we assume any or all of the following:

1. concentration of immobile factors
2. increasing returns to scale or indivisibility
3. externalities or public goods.

Cities arising from *concentrations of immobile factors* are relatively easily modeled, although we only mention them here. Given an immobile and concentrated factor, like a coal bed, industries which use the factor, such as mining, locate at that point. Industries such as steel, which use the primary product intensively, tend to locate nearby to save transportation costs. Others which are related and a retail sector follow for the same reason. The neoclassical model can describe such a city: there is convexity in production technology, and there are no externalities, and therefore the market mechanism can achieve an optimal allocation.[1] A concentration of immobile

[1] As discussed in Chapter I and Appendix I, there is a concealed nonconvexity in residential land use models, since a household can choose only one location. The nonconvexity, however, does not affect

factors, however, can produce only a relatively small city, and does not seem to be an important cause of modern cities.

We will first examine cities that arise from *economies of scale.* Economies of scale are prevalent in modern technology and result from such things as the division of labor and the indivisibility of such factor inputs as machinery and buildings. If the reduction in production costs due to scale economy is greater than the increase in commuting costs, a city will emerge. Such a city is basically a factory surrounded by the residential zone of its workers and may well be called a *'factory town'*.

Modern cities are, however, too complex to have resulted from simple economies of scale. Why should industries gather in a large city, where the commuting costs for each are greater than they would be in a single industry town? The answer is that industries find it profitable to gather together for a variety of reasons: communication costs and transportation costs of intermediate inputs can be saved; there is a larger pool of skilled labour to draw on, for example, and a more sophisticated infrastructure including transportation facilities. In order to capture these elements in a simple model, we assume a variant of *Marshallian externality.* We assume externalities among firms in a city, rather than among firms in an *industry*: all firms in a *city* are assumed to benefit from an increase in the population of the city. This assumption introduces the possibility of a city consisting of many firms by allowing increasing returns to scale which are internal to a city but external to the separate firms in a city.

We assume identical cities in both the increasing-returns-to-scale and the Marshallian-externality cases. This assumption allows us to obtain clear-cut results, but obviously fails to capture the complexity of the system of cities in the real world. In the last part of this chapter we discuss some possible extensions of the model, and the associated difficulties.

One of the major theoretical issues we try to analyze in this chapter is whether the decentralized market system can achieve the optimal allocation of cities, especially the optimal city size. If it cannot, we want to know how to correct the misallocation. It is well known in Welfare Economics that competitive equilibrium is not usually Pareto optimal in the presence of increasing returns to scale or externalities: there is almost always room to make somebody better off without making any others worse off. Although the result holds in our model, it is possible to describe an institutional arrangement that leads to optimal allocation.

The major difficulty in achieving an efficient allocation of an increasing-returns-to-scale industry is that the average cost always exceeds the marginal cost. Since an efficient allocation requires that the price be set equal to the marginal cost, the total revenue does not cover the total cost and the profit of a firm is negative. It is difficult to give such a firm a subsidy to cover the loss without destroying the incentive to minimize costs.

It turns out, however, that the loss of an urban producer equals the *aggregate*

---

the optimality of competitive equilibrium.

differential rent (the competitive urban rent minus the rural rent) of a city when the number of cities is optimal. This suggests that the optimal allocation could be achieved by a system of land developers. In each city an urban producer would lease all the land necessary for a city, including the residential area, from the rural landlords at the rural rent; sublease it to households at the competitive urban rent; and maximize differential rent plus profit. We will show that this arrangement does achieve the optimum allocation.

The optimal allocation of an economy with externalities requires *Pigouvian tax-subsidies*: agents that induce external costs or benefits for other agents must pay taxes or be given subsidies. In our model the population of a city gives external benefits to urban producers, so urban residents must be given subsidies. A relationship similar to that between the profit and the differential rent in the increasing returns to scale case holds with respect to the Pigouvian subsidy and the differential rent: at the optimum number of cities the aggregate Pigouvian subsidy equals the aggregate differential rent. This might seem to suggest that the optimal solution can be attained through the market if city governments return the differential rent to city residents as an equal social dividend. Unfortunately, this is not true. Though the optimal solution is indeed a market equilibrium under this institutional arrangement, city sizes greater than the optimum can also be equilibria.

1. The Model

The basic model of Chapter I becomes a simple general equilibrium framework when the *rural sector* is explicitly introduced. Consider a flat and fertile plain over which the rural sector is spread out. Circular cities are sprinkled about on the plain as in Figure 1. The plain is so large that the cities do not overlap.

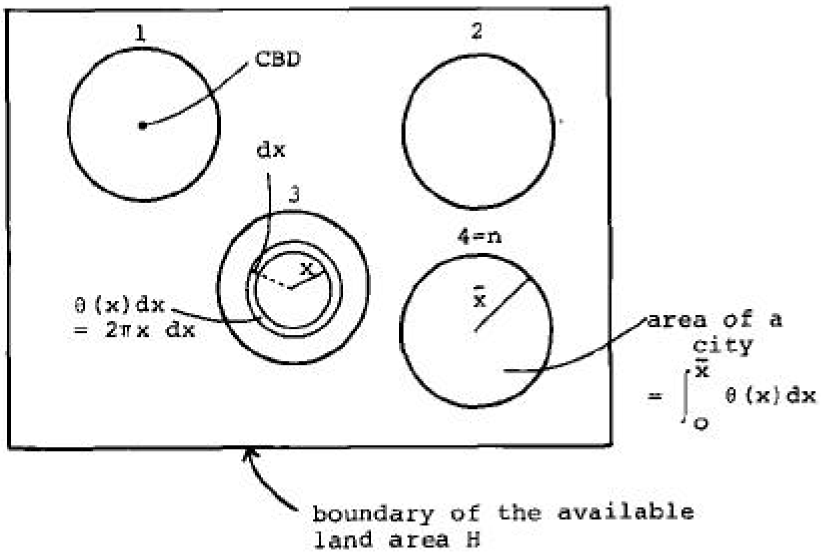

Figure 1. The Spatial Configuration of the Economy

We will imagine that each city consists of a business and production core surrounded by a residential zone. For the sake of simplicity, we will assume that urban production takes no space and no materials, so that only labor enters the production function. As before, households are identical, and each household has a single worker who commutes to a job at the center. We use the words 'households' and 'workers' inter-changeably, and the number of households is treated as the population. Commuting involves transportation costs but we assume that both urban and rural goods can be moved costlessly.

The production function of an urban firm is

$$f(\ell, P_c), \tag{1.1}$$

where ℓ and P_c are respectively labour input to a firm and the population of the city. We assume that all firms are identical. If the number of firms is m, the total production of a city is

$$Y = mf(\ell, P_c), \tag{1.2}$$

where

$$m = \frac{P_c}{\ell}. \tag{1.3}$$

Until section 5 we will assume that $f_p(\ell, P_c) = 0$ so that firms gain no advantage from increased population. Since grouping firms will result in higher commuting costs, there will be only one firm in a city. In this case the aggregate urban production function, $F(P_c)$ and an individual firm's production function will be identical:

$$F(P_c) \equiv f(P_c, P_c). \tag{1.4}$$

The production function is differentiable, with a positive marginal product. In order to allow for the possibility of increasing returns to scale, we do *not* impose the condition that the production function be concave.

We assume that the rural sector produces some complete and useful product, say soy beans, and that it has constant returns to scale. The aggregate production function of the rural sector is

$$G(P_a, H_a)), \tag{1.5}$$

where P_a and H_a are respectively the aggregate labour and land inputs. We assume that the production function is concave, homogeneous of degree one, differentiable, and that it has positive marginal products.

Goods produced by rural and urban sectors are called the *rural* and the *urban goods* respectively. Both goods are consumed by households. All households have the same differentiable, quasi-concave utility function,

$$u(a,z,h), \tag{1.6}$$

where a, z, and h are respectively the rural good, the urban good, and land for housing and where all goods are assumed to have positive marginal utilities.

The number of cities is denoted by n. For simplicity, we assume that all cities are identical, and we use the same notation for all cities. Since all cities have the same technology, and all households are the same, the assumption can usually be justified when the number of cities is large enough.

All cities are circular, and the amount of land available for housing between x and $x+dx$ from the center is $q(x)dx = 2pxdx$. As the production does not require land, the residential zone stretches from 0 to $\bar{x}$. The consumption of land for housing is constrained by

$$h(x)N(x) = q(x), \qquad 0 \le x \le \bar{x}, \tag{1.7}$$

where $N(x)dx$ is the number of households which live between x and $x+dx$, and $h(x)$ is the lot size of a house at x.

The total available land for the economy, H, is divided between cities and rural areas. The rural sector uses land both for production and for housing the rural workers. The land constraint for the entire economy is

$$H = n\int_0^{\bar{x}} q(x)dx + P_a h + H_a, \tag{1.8}$$

where h denotes the consumption of housing by rural residents. Note that, if h appears without the argument x, it denotes consumption by a rural resident. This distinction in notation will be used consistently. Implicit in constraint (1.8) is the assumption that the total available land is large enough to preclude overlapping of city areas.

Transportation requires many different inputs, but for the sake of simplicity, we assume that only the rural good is consumed in commuting. We continue to assume that goods, urban and rural, can be transported costlessly. The market clearing conditions for rural and urban goods are respectively

$$G(P_a, H_a) = n\int_0^{\bar{x}} [a(x) + t(x)]N(x)dx + P_a a, \tag{1.9}$$

$$nF(P_c) = n\int_0^{\bar{x}} z(x)N(x)dx + P_a z, \tag{1.10}$$

where $z(x)$ and $h(x)$ are the urban consumptions of urban and rural goods, and z and a are the rural consumptions. Commuting costs, $t(x)$, for a city resident at x satisfy

$$t(0) = 0\,. \tag{1.11}$$

The labour force in a city is assumed to equal the number of households living in the city:

$$P_c = \int_0^{\overline{x}} N(x)dx\,. \tag{1.12}$$

The population, P, of the whole economy, which is assumed to be given, is divided into urban and rural sectors:

$$P = nP_c + P_c\,. \tag{1.13}$$

2 A Fixed Number of Cities

We first derive the optimal solution under the assumption that the number of cities is exogenously given. As mentioned above, in this and the next two sections we assume that the marginal effect on production of increasing the population of a city is zero: $f_P(\ell, P_c) = 0$. Firms gain no advantage having other firms in the city, therefore, and each city has only one firm. The cities which we consider are based on economies of scale which are internal to the firm: $f_\ell > f/\ell$. Since $P_c = \ell$ for a single firm city, our notation can be simplified by using the aggregate production function, $F(P_c)$, and assuming

$$F'(P_c) > F(P_c)/P_c\,. \tag{2.1}$$

The utility level is maximized subject to the constraints (1.7)-(1.10), (1.12), (1.13), and the equal-utility constraints,

$$u(a(x), z(x), h(x)) = u\,,\quad 0 \le x \le \overline{x}\,, \tag{2.2}$$

$$u(a, z, h) = u\,, \tag{2.3}$$

which require that all households receive the same utility level.

The Lagrangian for this problem is

$$\begin{aligned}\Lambda = u &+ n\int_0^{\bar{x}} n(x)[u(a(x),z(x),h(x))-u]N(x)dx + n_a[u(a,z,h)-u]P_a \\ &+ d_c[nF(\int_0^{\bar{x}} N(x)dx) - n\int_0^{\bar{x}} z(x)N(x)dx - P_a z] \\ &+ d_a\left\{G(P_a,H_a) - n\int_0^{\bar{x}}[a(x)+t(x)]N(x)dx - P_a a\right\} \\ &+ n\int_0^{\bar{x}} m(x)[q(x)-h(x)N(x)]dx \\ &+ m[H - n\int_0^{\bar{x}} q(x)dx - P_a h - H_a] \\ &+ g\left[P - n\int_0^{\bar{x}} N(x)dx - P_a\right]\end{aligned} \tag{2.4}$$

where control variables are $a(x), z(x), h(x)$, and $N(x)$; control parameters are a, z, h, u, P_a, H_a, and $\bar{x}$; and $n(x)$, n_a, d_c, d_a, $m(x)$, m, and g are respectively Lagrange multipliers for (2.2), (2.3), (1.10), (1.9), (1.7), (1.8), and (1.13). Note that P_c is eliminated by substituting (1.12) into (1.10) and (1.13). The Lagrange multipliers have basically the same interpretation as in Chapter I: $n(x)$ and n_a are weights attached to the utilities of different households to obtain equal utility levels; d_c and d_a are respectively shadow prices of the urban and rural goods; $m(x)$ is the shadow rent of land at distance x from the center of a city and m the shadow rent of the rural land; and g is the shadow 'price' of a household.

The Lagrange multipliers express shadow prices in utility terms. It is convenient to transform shadow prices into pecuniary terms. Taking the rural good as a numeraire, we define $p = d_c / d_a$, $R(x) = m(x)/d_a$, $R_a = m/d_a$ and $s = -g/d_a$. p is the shadow price of the urban good, $R(x)$ the shadow rent at x, R_a the shadow rent of the rural land, and a the marginal social cost—the negative of the shadow price—of a household.

First order conditions are immediately obtained by applying the result in the appendix on optimum control theory. They become, after simple rearrangements,

$$\frac{u_z(a(x),z(x),h(x))}{u_a(a(x),z(x),h(x))} = \frac{u_z(a,z,h)}{u_a(a,z,h)} = p, \qquad 0 \le x \le \bar{x}, \tag{2.5a}$$

$$\frac{u_h(a(x),z(x),h(x))}{u_a(a(x),z(x),h(x))} = R(x), \qquad 0 \le x \le \bar{x}, \tag{2.5b}$$

$$\frac{u_h(a,z,h)}{u_a(a,z,h)} = R_a, \tag{2.5c}$$

$$G_H = R_a, \tag{2.5d}$$

$$G_P + s = a + pz + R_a h\,, \tag{2.5e}$$

$$pF' + s = a(x) + pz(x) + R(x)h(x) + t(x)\,, \qquad 0 \le x \le \bar{x}\,, \tag{2.5f}$$

$$R(\bar{x}) = R_a\,.$$

(2.5a)-(2.5c) are the usual conditions equating marginal rates of substitution to (shadow) prices. Note that (2.5a) implies that all households have the same marginal rates of substitution between the urban good and the rural good. This is an immediate consequence of our assumption that it costs nothing to transport either good.

(2.5d) states that the value of marginal productivity of land is equal to the shadow rent of land. From (2.5c), (2.5d) and (2.5g), rural households, rural producers and urban residents at the edge of a city all face the same shadow rent. This condition implies that shadow rent varies continuously over space.

The social cost, s, of a household must be the value of resources it consumes minus the value of its marginal product:

$$s = a + pz + R_a h - G_p$$

$$s = a(x) + pz(x) + R(x)h(x) + t(x) - pF'\,.$$

These yield respectively (2.5e) and (2.5f).

If workers are paid the value of their marginal products, and if all prices equal shadow prices, a household must be given a subsidy which is equal to s in order to satisfy the budget constraint. Because of the resource constraints, (1.9) and (1.10), however, the sum of the subsidies must equal the total surplus in the economy, which is the sum of the total rent and the total profit:

$$s = \frac{l}{P}\left\{ n\int_0^{\bar{x}} R(x)q(x)dx + R_a\left[P_a h + H_a\right] + np\left[F + P_c F'\right]\right\}. \tag{2.6}$$

If this optimal solution is decentralized using the usual price mechanism, an urban producer might incur a loss at the optimum, since we allowed increasing returns to scale. In such a case the government must give the producer a subsidy equal to the loss. If the subsidy does not weaken a firm's incentive to minimize costs, the price mechanism attains the optimal allocation. Unfortunately, administering such a subsidy requires a prohibitive amount of information.

These problems may not arise if the producer can act as land developer, collecting the residential land rent. We consider this institutional framework in section 4.

Of course, if the urban sector has decreasing or constant returns to scale, the urban sector earns a positive or zero profit, and hence a subsidy is not necessary. In such cases, however, there is no reason to have cities, since by reducing city size transportation costs can always be reduced without raising production costs. As will be shown in the next section, the optimal solution with nonincreasing returns to scale requires that the urban sector spread uniformly over space.

3. A Variable Number of Cities

When we treat the number of cities as an endogenous variable, it is convenient to assume that the number is so large that it can be safely approximated by a continuous variable. Taking a derivative of the Lagrangian (2.4) with respect to n and substituting other first order conditions, we obtain

$$pF(P_c) - [pP_c F'(P_c) - \int_0^{\overline{x}} R(x) q(x) dx] - R_a \int_0^{\overline{x}} q(x) dx = 0. \tag{3.1}$$

The number of cities should be increased up to the point where an additional city has zero social net value. The net value of an additional city is the value of the gross product of the city minus the costs of producing it. The value of the gross product is pF. The cost of producing it is the cost of supporting workers. Workers consume the rural good, the urban good, and land: and they pay commuting costs. The total social value of their consumption is

$$\int_0^{\overline{x}} [a(x) + pz(x) + t(x)] N(x) dx + R_a \int_0^{\overline{x}} q(x) dx.,$$

where land must be evaluated at the opportunity cost, R_a, instead of the urban rent. This is not yet the social cost of supporting workers of an additional city. Since the society incurs the social cost of a household, *s*, regardless of whether a household lives in the city or not, sP_c must be subtracted from the costs. The net value of an additional city is then

$$pF(P_c) - \int_0^{\overline{x}} [a(x) + pz(x) + t(x)] N(x) dx - R_a \int_0^{\overline{x}} q(x) dx + sP_c = 0.$$

Substitution of (2.5f) into this equation yields (3.1).

If the number of cities is optimal, the population of a city is such that it minimizes the value of per capita consumption of resources, or the average cost of maintaining the utility level. Otherwise, there is some other population level which achieves the same utility level with a lower per capita consumption of resources, and the value of resources used by the entire urban sector can be reduced by changing the population size of all cities while changing the number of cities accordingly to keep the population of the entire urban sector unchanged. The resources saved could be used to raise the utility level, which proves that the allocation cannot be optimal. Note that consumption of resources in this argument does not need to be modified by subtracting the social cost of a household.

This observation facilitates another interesting interpretation of the optimality condition.[2] The net cost of maintaining the utility level is equal to the total city consumption, plus the opportunity cost of land, minus total production. Expressed in terms of city population, the net cost is simply a total cost function,

$$TC(P_c) = \int_0^{\bar{x}} [a(x) + pz(x) + t(x)]N(x)dx + R_a \int_0^{\bar{x}} q(x)dx - pF(P_c) \quad .$$

The per capita, or average, cost is

$$AC(P_c) = TC(P_c)/P_c \; .$$

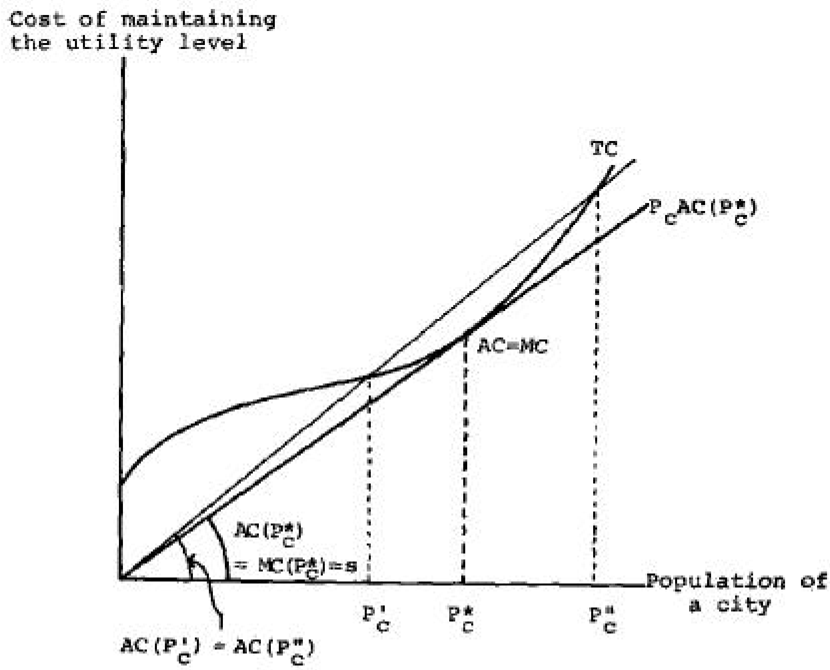

Figure 2. Optimum City Size

As illustrated in Figure 2, the average cost is minimized when it equals the marginal cost. The marginal cost is the cost of adding a household to the city. Since at the optimum the cost of adding a household must equal everywhere in the city, we may consider the addition at any radius. Addition at the edge of the city is easiest because there is no ambiguity about whether the rural rent or the urban rent expresses the value of land. The cost of adding a household at $\bar{x}$ is the value of consumption minus the marginal product:

$$MC(P_c) = a(\bar{x}) + pz(\bar{x}) + t(\bar{x}) + R_a h(\bar{x}) - pF'(P_c) \quad ,$$

which, from (2.5f), equals the social cost of a household, s, and also equals the value of

[2] This interpretation was suggested by Arnott. Similar interpretation is published in Arnott (1979).

consumption minus the marginal product at any other radius when consumption of land is evaluated at the urban shadow rent:

$$MC(P_c) = a(x) + pz(x) + t(x) + R(x)h(x) - pF'(P_c) \quad , \qquad 0 \le x \le \bar{x} \quad .$$

Multiplying this equation by $N(x)$ and integrating it from 0 to $\bar{x}$ yields

$$P_c MC(P_c) = \int_0^{\bar{x}} [a(x) + pz(x) + t(x)]N(x)dx + \int_0^{\bar{x}} R(x)q(x)dx - pP_c F'(P_c) \quad .$$

At the optimum the average cost equals the marginal cost, which implies equation (3.1) above.

We can rewrite (3.1) as

$$-p[F(P_c) - P_c F'(P_c)] = \int_0^{\bar{x}} [R(x) - R_a]q(x)dx \quad . \qquad (3.1')$$

If factor prices are equal to the values of marginal productivities, the left can be interpreted as the operating loss of an urban firm. Then (3.1') states that, *in a single firm city the firm's operating loss is equal to the aggregate differential rent* (the urban rent minus the rural rent) *if the number of cities is optimal.* Using this equation, we can rewrite (2.6) as

$$s = HR_a / P , \qquad (3.2)$$

which says that the social cost of a household equals the per capita rural rent.

With constant or decreasing returns to scale, firms earn nonnegative profits. By (3.1') the aggregate differential rent would be nonpositive at the optimum, implying that cities have simply vanished. This is quite reasonable since smaller cities have the advantage of lower transportation costs with no disadvantage on the production side.

If the urban sector has increasing returns to scale, bigger cities have the advantage of lower average production costs. The optimum city size or the optimum number of cities is determined so as to balance the transportation costs and the benefit from increasing returns to scale. (3.1) shows that this balance is attained when the loss of the urban sector equals the aggregate differential rent.

By solving the problem of Section 2, the utility level can be obtained as a function, u(n), of the number of cities. The second order condition requires that

$$\frac{d^2u(n)}{dn^2} \le 0 \, .$$

By the Envelope Theorem,[3] $du(n)/dn$ is equal to the partial derivative of the Lagrangian (2.4) with respect to *n*.

$$\frac{du(n)}{dn} = \frac{\partial \Lambda}{\partial n} = d_a\{p[F - P_c F'] + \int_0^{\bar{x}} [R(x) - R_a] q(x) dx\} \tag{3.3}$$

Hence, using the fact that the term with dd_a/dn vanishes by (3.1), the second derivative is

$$\frac{d^2 u(n)}{dn^2} = d_a \frac{d}{dn}\{p[F - P_c F'] + \int_0^{\bar{x}} [R(x) - R_a] q(x) dx\}, \tag{3.4}$$

which must be nonpositive at a maximum.

The net benefit from an additional city is the sum of differential rent and profit. The optimum is attained at the point where the net benefit is zero. In order to have maximum rather than minimum, however, the net benefit must be decreasing at the optimum, and the sum of the aggregate differential rent and the profit of an urban producer must be a decreasing function of the number of cities.

The second order condition is usually satisfied if the degree of increasing returns to scale declines as a city becomes larger, because the aggregate differential rent is usually larger in a larger city.

An important implication of this second order condition is that, if the aggregate differential rent exceeds the loss of the urban sector, there are too few cities. Since cities tend to be larger when there are fewer cities, city size is likely to be too big in this case. Notice, however, that this result is obtained under the condition that all variables other than the number of cities are optimally chosen. If there is some distortion like monopsony pricing in the labour market, this condition is not satisfied, and the difference between the differential rent and a firm's loss does not serve as a signal of whether or not the city size is too big. In the next section a paradoxical kind of monopsony pricing will be shown to exist at the market equilibrium of our model.

4. Market City Sizes

As noted in section 2, a firm must be given an appropriate subsidy to achieve the optimal solution in a decentralized market system. The result in section 3 shows that if the number of cities is optimal, the subsidy equals the total differential rent of a city. Thus the optimal solution can be decentralized by giving a firm a subsidy equal to the differential rent and distributing among all rural and urban households the rest of the rent (which equals the rural rent for the entire land, $R_a H$) as an equal lump-sum social dividend.

---

[3] See Appendix III for the explanation of the Envelope Theorem.

There is, however, a way to achieve the optimal solution that does not require as much knowledge and action on the part of the government. It turns out that *the optimal solution can be achieved by allowing firms to lease the urban land including the residential area.* The entire available land is owned collectively by all households in the economy. Rural producers and rural households rent the land and pay the rural rent. An urban firm, acting as a developer of a whole city, also rents urban land at the rural rent, but subleases it to city residents at the competitive market rent. Then the firm maximizes the sum of profit from production, which is usually negative, and the net rent. Firms, like land, are owned collectively by all households, and profits are distributed equally among all households. We show that such a system of urban-producers/ city-developers attains the optimal allocation, providing, of course, that the firms are perfectly competitive.

The number of cities (and hence the number of firms) is assumed to be so large that a firm acts as a price and utility taker: since there are no transportation costs for the urban good, firms directly compete with each other in the product market, and a single firm cannot significantly affect the price of the urban good. Under the assumption of perfect mobility, households move to the city where they can obtain the highest level of utility. Faced with freely mobile households, a firm must make sure that its employees obtain at least the same utility level as they would in any other city. This leads to utility taking behaviour as the number of firms becomes large. Notice, however, that a firm does not take the wage rate as given. Households decide to migrate on the basis of the utility level and not the wage rate. As long as the utility level is not lower than at any other place, the wage rate can be freely chosen.

A firm maximizes the profit on the entire development which is the revenue from the sales of its product, plus the land rent, minus the total wage bill, minus the total payment of the rural rent:

$$pF(P_c) + \int_0^{\bar{x}} R(x)q(x)dx - wP_c - R_a \int_0^{\bar{x}} q(x)dx \,, \tag{4.1}$$

where w is the wage rate.

Four variables, w, P_c, $\bar{x}$, and $R(x)$, are involved in this maximization problem, but the firm faces the constraints imposed by competition with other firms. The maximum rent that households can pay, if they are to achieve the given utility level, is a function of their wage. We can, therefore, reduce the problem to that of maximizing (4.1) with respect to the wage. First, using the indirect utility function of households, we express $R(x)$ as a function of the wage.

Since all firms and the entire land are collectively owned by all households in the economy, households obtain equal shares of profits of firms and the revenue from the rural rent paid by both the rural and urban sectors. Then the budget constraint is

$$w + s = a(x) + pz(x) + R(x)h(x) + t(x) \quad , \qquad 0 \le x \le \bar{x} \,, \tag{4.2}$$

where s is the share of the rent and profit, and satisfies (2.6). The following *indirect* *utility function* is obtained as a result of utility maximization under the budget constraint:

$$v(I(x), p, R(x)), \tag{4.3}$$

where $I(x)$ is the net income:

$$I(x) \equiv w + s - t(x). \tag{4.4}$$

Since the utility level, u, is taken as given, a firm maximizes profit under the constraint:

$$v(I(x), p, R(x)) = u. \tag{4.5}$$

This constraint enables us to express $R(x)$ as the *bid rent function,*

$$R(x) = R(I(x), p, u). \tag{4.6}$$

As in (I.1.14), we have

$$R_I(I(x), p, u) = 1/h(x), \tag{4.7}$$

where R_I is the partial derivative of the bid rent function with respect to the net income $I(x)$.

Substituting (4.7) and (4.4) into (4.1), we see that the firm's problem is to maximize

$$pF(P_c) - wP_c - R_a \int_0^{\bar{x}} q(x)dx + \int_0^{\bar{x}} R(w + s - t(x), p, u)q(x)dx \tag{4.8}$$

subject to the constraints

$$P_c = \int_0^{\bar{x}} R_I(w + s - t(x), p, u)q(x)dx, \tag{4.9}$$

$$R(w + s - t(\bar{x}), p, u) = R_a, \tag{4.10}$$

where p, u, and s are taken as given since the firm is small.

Now the population of the city can also be written as a function of the wage. Although the price of the product and the utility level are taken as fixed, the wage rate affects the supply of labour, because households will move to achieve the given utility level. From (4.10), $\bar{x}$ can be expressed as a function of w, and (4.9) becomes the following labour supply function,

$$P_c(w) = \int_0^{\bar{x}(w)} R_I(w + s - t(x), p, u)q(x)dx, \tag{4.11}$$

where s is given by (2.6).

Using this labour supply function we can demonstrate that firms have a kind of monopsony power despite the fact that they are competitive in the usual sense. The slope of the

supply curve is

$$P_c'(w) = \int_0^{\bar{x}} R_{II} q(x)dx + R_I(I(\bar{x}), p, u)q(\bar{x})\bar{x}'(w)$$

$$= \int_0^{\bar{x}} R_{II} q(x)dx + N(\bar{x})/t'(\bar{x}), \qquad (4.12)$$

where the second equality is obtained from (4.7) and

$$\bar{x}'(w) = 1/t'(\bar{x}). \qquad (4.13)$$

Since R_I satisfies

$$R_I(I(x), p, u) = 1/h[R(I(x), p, u)p, u], \qquad (4.14)$$

where $h[\ ]$ is a compensated demand function for land, we have

$$R_{II}(I(x), p, u) = -h_R R_I / h^2 = -h_R / h^3 > 0. \qquad (4.15)$$

Hence, (4.12) becomes

$$P_c'(w) = \int_0^{\bar{x}} -(h_R / h^2)N(x)dx + N(\bar{x})/t'(\bar{x}) > 0. \qquad (4.16)$$

Thus *the supply curve of labour is upward sloping, and firms have apparent monopsony power in the labour market.*

Using the labour supply function (4.11), we can also reduce the problem (4.8)-(4.10) to maximization of

$$pF(P_c(w)) - wP_c(w) - R_a \int_0^{\bar{x}(w)} q(x)dx + \int_0^{\bar{x}(w)} R(w + s - t(x), p, u)q(x)dx \qquad (4.17)$$

with respect to *w*. The first order condition is

$$(pF' - w)P_c'(w) - P_c(w) + \int_0^{\bar{x}} R_I q(x)dx + [R(w + s - t(\bar{x}), p, u) - R_a]q(\bar{x})\bar{x}'(w) = 0$$

(4.18)

From (4.10) and (4.11), this becomes

$$pF'(P_c) = w. \qquad (4.19)$$

Thus, even though a firm faces an upward-sloping supply curve in the labour market, it behaves like a price taker and sets the wage rate at the level where the value of marginal product equals the wage.

The reason is that when the utility level is fixed, the decrease in income of workers is fully reflected in a decrease in expenditures on land. Thus the increase in profit caused by lowering the wage is completely offset by the decrease in land rent, and the firm behaves as if there were no monopsony gain.

It now follows that all the first order conditions for the optimum are satisfied in market equilibrium: (2.5f) is obtained from (4.2) and (4.19); (2.5g) is equivalent to (4.10); (2.5a)-(2.5c) are the results of utility maximization of households; (2.5d) and (2.5e) result from profit maximization in the rural sector; and free entry insures equation (3.1), which states that the maximized profit (including differential rent) is zero in equilibrium. Therefore, if the first order conditions are sufficient to characterize the optimal solution, *the market equilibrium is optimal under the institutional framework in which a firm can act as the developer of an entire city.*

When firms cannot act as developers, however, a market equilibrium differs from the optimal allocation. A firm maximizes

$$pF(P_c) - wP_c, \tag{4.20}$$

taking the price of the product and the utility level of the workers as given. Households receive equal shares of profits of firms and the total land rent including the urban rent, and s is given by (2.6). In this case, too, a firm has monopsony power in the labour market in the sense that it faces an upward sloping supply curve. The labour supply function is given by (4.11). A firm takes s as given in this case, as in the last.

The first order condition for profit maximization is

$$pF'(P_c) = w + P_c / P_c'(w). \tag{4.21}$$

Free entry insures that the maximized profit is zero:

$$pF(P_c) - wP_c = 0. \tag{4.22}$$

Multiplying (4.21) by P_c and subtracting it from (4.22), we obtain

$$p[F(P_c) - P_c F'(P_c)] = -P_c^2 / P_c'(w) < 0, \tag{4.23}$$

which implies that market equilibrium occurs when the firm operates in the region of increasing returns to scale. The profit is, however, zero, since by (4.21) a firm exploits monopsony power, and pays a wage rate lower than the value of the marginal product of labour.

From (3.1), the optimal city size also occurs in the region of increasing returns to scale. Since market equilibrium and optimal city sizes are in the same range of the production function, observation about returns to scale cannot be used to determine which is larger in general. It depends on the amount of monopsony power and the size of the aggregate differential rent. In principle there is no reason why the equilibrium city size should coincide with optimum city size, and a city in which firms maximize profits but do not act as land developers has zero probability of achieving an exactly optimum city size.

5. The Marshallian Externality Case

In the previous section scale economies were internal to the firm. We now assume scale economy internal to a city but external to a firm. Cities form because firms are more productive if they can draw on a larger population. To capture this effect, we repeat the analysis of the previous sections with one change. The production function of a firm is still

$$f(\ell, P_c) \tag{1.1}$$

but now we allow an increase in population to increase the firm's productivity:

$$f_P > 0\,. \tag{5.1}$$

This version of the *Marshallian externality* can result in multi-firm cities, since the presence of additional firms is now an advantage. If m is the number of firms, the total product of a city is

$$Y = mf(\ell, P_c) \tag{1.2}$$

where

$$m = \frac{P_c}{\ell}\,. \tag{1.3}$$

We assume increasing average returns to labour when the firm is small, with a gradual shift to decreasing returns as ℓ increases.

As in the previous case, we first consider the case of a L, fixed number of cities. The optimization problem can be solved in the same way as before, if (1.2) and (1.3) are substituted into the proper places. The first order condition (2.5f) is replaced by two conditions:

$$f = \ell f_\ell\,, \tag{5.2}$$

$$pf_\ell + s_c + s = a(x) + pz(x) + R(x)h(x) + t(x)\,, \quad 0 \le x \le \bar{x}\,, \tag{5.3}$$

where

$$s_c = pmf_P \tag{5.4}$$

(5.2) shows that the marginal product of labour equals the average product, and implies that firms operate under constant returns to scale at the optimum. The optimum is attained in long-run equilibrium at which all firms operate at the bottom of the average cost curve since constant returns to scale hold when the average cost curve is flat.

From (5.3), the total value of household consumption equals the sum of the three terms on the left: the value of marginal productivity of labour; the value of the marginal external economy, s_c, which an urban resident gives to the urban production sector; and the social dividend, s, equal for all households in both cities and the rural area. Therefore, a household must be given the Pigouvian subsidy, s_c, in addition to wages and the social dividend, *s*. A city resident gives external benefits to urban producers, and should be given a subsidy equal to the value of his marginal contribution to urban production.

Taking the number of cities as a variable now, the optimality condition becomes, after simple rearrangements,

$$s_c P_c = \int_0^{\bar{x}} [R(x) - R_a] q(x) dx . \tag{5.5}$$

Thus, at the optimum the total Pigouvian subsidy equals the total differential rent in a city. As before, the equal lump-sum social dividend s is given by (3.2).[4]

The optimal allocation is a market equilibrium in the following institutional setting. All land is equally and collectively owned by all households in the economy. Residents in a city form a cooperative, or a city government, which rents all the land for the city at the rural rent. Each household, in turn, rents land for housing from the city government, and pays the market-determined rent. Since the urban residential rent is higher than the rural rent, the city government has a surplus revenue. The surplus is returned to city residents as an equal subsidy. It will be shown later that the optimal

---

[4] It is easy to show that (5.5) is equivalent to (3.1). From (5.2), we can express labour input to a firm as a function, $\ell(P_c)$, of the population of a city. The aggregate production function can then be written as

$$F(P_c) = \frac{P_c}{\ell(P_c)} f(\ell(P_c), P_c).$$

Differentiation of the aggregate production function yields

$$\begin{aligned} F'(P_c) &= \frac{P_c}{\ell} f_P(\ell, P_c) + \frac{1}{\ell} f(\ell, P_c) + \left[\frac{P_c}{\ell} f_\ell(\ell, P_c) - \frac{P_c}{\ell^2} f(\ell, P_c) \right] \ell'(P_c) \\ &= \frac{P_c}{\ell} f_P + \frac{1}{\ell} f \end{aligned}$$

where the second equality is obtained from(5.2). Noting

(1.3) and (5.4), we finally obtain

$$s_c P_c = -p[F(P_c) - P_c F'(P_c)]$$

which shows that (5.5) is equivalent to (3.1).

solution is a market equilibrium under this institutional arrangement. Unfortunately, however, the optimal allocation is not a *unique* market equilibrium. A wide range of city sizes greater than the optimum can also be equilibria, and there is no reason to believe that the optimum is likely to be attained.

This point can be illustrated in the following way. If we specify the number of cities, a market equilibrium is obtained by substituting

$$s_c = \frac{1}{P_c}\int_0^{\bar{x}} \left[R(x) - R_a\right] q(x)dx$$

and (3.2) into the first order conditions (2.5a)-(2.5e), (2.5g), (5.2) and (5.3). Since the resulting population size is normally the same for all cities, we can consider the equilibrium utility level as a function, $u^*(P_c)$, of the population of a city. For simplicity, the function is assumed to be single-peaked as in Figure 3.

Clearly, city sizes less than P_c^*, where the equilibrium utility level attains its maximum, cannot be equilibria. If a household moves to another city, the utility level will rise in the receiving city, and fall in the city which has lost population. Therefore, a household has an incentive to move to another city. The receiving city would continue to grow at least until P_c^* was reached. The losing city would eventually disappear.

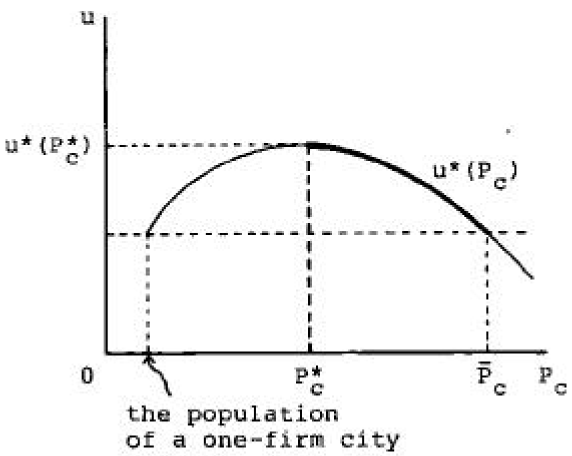

Figure 3. Market City Sizes

City sizes greater than P_c^*, however, can be an equilibrium. Households do not have an incentive to move to another city at a city size greater than P_c^* since an increase in the population of the receiving city would lower the utility there. For the same reason they do not have an incentive to move to the rural area, either. The only way, therefore, to reduce the city size is to create a new city. If we do not allow for coalition or entrepreneurship to form a large new city, all new cities must start from one firm. In this case, a new city will not be formed unless the size of existing cities exceeds P_c^* by so much that the general utility level falls to that of a one-firm city. Therefore, the city sizes between P_c^* and $\overline{P}_c$ tend to remain the same.

It is easy to see that the city size, P_c^*, which maximizes the utility level among market equilibria coincides with that of the direct optimum. Though the direct optimum does not have, among its constraints, the conditions for a market equilibrium, the first order conditions for the direct optimum include all the conditions required for a market equilibrium. The two problems, therefore, must have the same optimum.

Thus *the minimum market city size coincides with the optimal city size when there is a Marshallian externality, and there is a strong tendency for market city sizes to become too large.* This result suggests that government intervention is necessary to achieve the optimum city size

Whether intervention is required to not, the actual situation may be less serious than the model suggests. Historical development has provided us with a hierarchy of cities rather than a single type. Cities produce different sets of commodities, and bigger cities produce more commodities than smaller ones. A new city at a certain level of the hierarchy can be created by adding firms producing new commodities to an existing city at a lower level of the hierarchy. This does not require a very large population shift.

The above special institutional arrangement which allows cities to collect land rent and distribute the revenue among city residents is not usually possible in a private ownership economy. In a private ownership economy, migrational decisions are not affected by the land rent households can *earn*, though they are certainly affected by the level of the rent they must pay. One reason is that households may be able to invest in houses in cities where they do not live. Another is that even if all houses are owner-occupied, households must pay the discounted value of future rent when they move to a city. The benefit of future high rent, which might attract households to a city, is thereby neutralized by the purchase cost of the house. Thus the usual private ownership economy is closer to the case of $s_c = 0$ and the city size which maximizes the utility level among market equilibria is different from the direct optimum.

It is not clear in a general case whether this city size is bigger or smaller than the optimum city size. If there is no rural sector, it is obvious that this city size coincides with the optimum. Divergence from the optimum is caused by the fact that the absence of the Pigouvian subsidy distorts the allocation of households between the urban and the rural sectors. In the real world it seems likely that the population in the urban sector is too small, because the incentive to live in cities is weaker due to the lack of the Pigouvian subsidy. However, the problem is more subtle than it appears, since it involves determining the number of cities. It is not quite clear how the number of cities is affected by the distortion.

We have assumed that it is the population of a city that generates an external economy. Obviously, this is not the only formulation. For example, we can assume that the total product of a city induces the externality, as in Henderson (1974). In that case, the Pigouvian subsidy should be given to firms as excise subsidy on their products. With this change, the above analysis can be applied and the same conclusions are obtained.

6. Differences in City Sizes

So far we have considered only cities which have the same allocation, both at the optimum and in market equilibrium. This is clearly unrealistic. Relaxing the simplifying assumptions of previous sections, we can obtain differences among cities.

First, production functions may differ among cities because of differences in climate, factor endowment, and so on, or simply because technology does not diffuse instantaneously. Since cities with technological advantage tend to attract more households than others, city sizes vary.

This extension turns out to be fairly simple. In the case of increasing returns to scale, internal to a firm, the only change is that we must distinguish cities notationally since they in general have different allocations. The first order conditions (2.5a), (2.5b), (2.5f) and (2.5g) hold in all cities. In particular, (2.5f) and (2.5g) for the i-th city must read

$$pF_P^i(P_c^i, H_c^i) + s = a^i(x) + pz^i(x) + R^i(x)h^i(x) + t(x) \tag{6.1}$$

$$R^i(\bar{x}^i) = R_a . \tag{6.2}$$

Combining these equations, we have

$$pF_P^i + s - t(\bar{x}^i) = a^i(\bar{x}^i) + pz^i(\bar{x}^i) + R_a h^i(\bar{x}^i) . \tag{6.3}$$

Since all households must receive the same utility, the right side is equal for all cities. Hence, the value of the marginal product of labour minus the commuting costs at the edge of a city is the same for all cities:

$$pF_P^i - t(\bar{x}^i) = pF_P^j - t(\bar{x}^j) \qquad \text{for any } i, j. \tag{6.4}$$

When the number of firms is optimized, the *marginal* firm will obtain zero profit (including the aggregate differential rent) and other firms will earn positive profits.[5] In exactly the same way as in the case of identical cities, it can be shown that, if firms

[5] Here, it is implicitly assumed that there is no competitive bidding for the right to build a plant in a specific city. This is the reason why a firm located in an advantageous city earns excess profit. The profit is caused by the Presence of some unpriced factors such as good climate, clean water, etc. If these factors are competitively priced, all firms earn zero profit. Even if there is no market for these factors, competitive bidding for the site of a plant drives down the profit to zero and the rent is captured by the owner of the site.

act as land developers, the market equilibrium attains the optimal allocation.

We can analyze the Marshallian externality case in a similar way. It is easy to see that the value of the marginal product of labour, plus the Pigouvian tax, minus the commuting costs at the edge of a city is equal for all cities;

$$pf_\ell^i + s_c^i - t(\overline{x}^i) = pf_\ell^j + s_c^j - t(\overline{x}^j) \qquad \text{for any } i, j .$$

The condition for the optimum number of cities is that the aggregate Pigouvian subsidy equals the aggregate differential rent in the *marginal* city. However, the equality does not hold in inframarginal cities. This causes a difference from the case of identical cities. If all cities are identical, the aggregate Pigouvian subsidy must equal the aggregate differential rent in all cities. This is the reason why we obtained the result that, if the differential rent is returned to city residents as an equal subsidy, the optimal allocation is one of market equilibria. If cities are not identical, the result does not hold, since the optimum Pigouvian subsidy is not equal to the average differential rent in inframarginal cities. Therefore, even the best allocation among market equilibria does not coincide with the optimal allocation.

A second class of differences which can give rise to differing cities includes all the ways that household tastes and skills may vary. An extended analysis, unfortunately, is so messy that we have reluctantly decided to spare our readers.

Although it is certainly more realistic to include these factors, they alone cannot explain the differences we observe in modern economies. The fact that cities produce differing bundles of commodities probably explains more of the variation in their sizes than, for example, consumer tastes.

Consider the effect of introducing more than one urban good into the model with increasing returns to scale. If the goods have different production functions, the cities will have different sizes.

If we ask whether a city can produce more than one good in our model, we discover an important implication of the assumption that transporting goods costs nothing. Commuting costs can be saved by separating firms producing different goods, without incurring any additional costs, so two-product cities will not occur.

If transporting urban goods is costly, however, cities producing more than one good might well arise. The saving on transporting wet concrete or bottled coke to demanders, for example, might justify the extra commuter costs that result from having a concrete plant and a bottling plant in each city.

The cost of goods transport has a strong influence on city form as well as size, although the subject is outside the range of this chapter. Even if two or more commodities are produced in a city, the firms will not necessarily all locate at the center. Retail stores, for example, disperse throughout a city to reduce the transportation costs of shopping for consumers. Moreover, there is no a priori reason to expect that a concrete plant and a bottling plant locate at the center. They might locate at the edge of a city to take advantage of lower land costs, and form a multi-centered city.

There is another problem in multi-product cities caused specifically by the cost of goods transport. There can be only one firm with the greatest returns to scale in a city. If there were two, and if we could ignore the fact that the numbers of firms in other industries must be integers,[6] then we could split the city into two. Production costs would not increase in any industry and commuting costs would decrease, and society be better off with two cities instead of one.

Therefore, we have to introduce externalities in order to attain a more realistic system of cities. The simplest way is to add another urban good to the framework in section 5. If there are two urban goods, we obtain three types of cities:

two producing only one good, and one producing both. It is easy to see that the same results as those in section 5 can be obtained for each type of city.

However, there is no guarantee that cities producing both goods are bigger than cities producing only one good. For example, if externality works only through the total population of a city, cities producing two goods have no more benefit from becoming bigger than cities producing one good. Therefore, we might want to assume that there is a special benefit which arises from having two industries together.

Although introducing cross product externalities is attractive, and would give rise to more realistic system of cities in our model, the analysis is simply too difficult for the present work. We do not, therefore attempt to build a model of a system of cities of this type here.

Notes

Until Alonso's work (1971), the analysis of city sizes had been limited to the cost side, and the city size which gave the minimum cost had been considered optimal. Alonso introduced the output side, regarding a city as an aggregate production unit. There are two types of optimum city size in this model. For the residents the optimum size is that which maximizes the difference between the average product (AP) and the average cost (AC). For a national government interested in maximizing total product under conditions of labour surplus, the optimum size is where the marginal product (MP) is equal to the marginal cost (MC). If the supply of labour is limited, this condition should be modified. MP may not equal MC although the difference between MP and MC must be the same for all cities. Alonso pointed out that if individuals maximize the difference between AP and AC, per capita tax of MP-AP-(MC-AC) can result in the optimum city size.

Although Alonso's work was a big step forward in constructing the economic theory of city sizes, his approach has the following shortcomings. First, the analysis is partial in nature, since only one city is considered: if the city is placed in a general

[6] For example, if there are two firms of the greatest degree of increasing returns and three firms of the second greatest degree of increasing returns, splitting this city into two may involve an extra social cost since a city cannot have one and a half firms.

equilibrium framework, we may face different problems. Second, the welfare aspect of the analysis is not very clear, since utility functions for households are not introduced. Third, the cause of increasing average product is not explicitly formulated. It is not clear, therefore, how individual firms and households behave in a market economy: increasing returns to scale for a firm, and external economies among firms have very different implications on individual behaviour. Fourth, the spatial aspect of cities is ignored.

There have been several attempts to overcome these shortcomings. Borukhov (1975) built a very simple model of an economy consisting of many cities. He showed that Alonso's second condition for the optimum city size is correct if the number of cities is given: at the optimum the difference between MP and MC is equal for all cities, but MP exceeds MC by an amount which has been interpreted as the opportunity costs of siting the population in alternative cities. If the number of cities is a variable, however, this condition is not sufficient to characterize the optimal solution. Since Borukhov was worried about integerness of the number of cities, he could not obtain a transparent condition for the optimum number of cities. However, if one is willing to approximate the number of cities by a continuous variable, and to assume that all cities are the same (as done in this chapter), it is easy to see that at the optimum the difference between MP and MC is equal to the difference between AP and AC. This means that the difference between AP and AC is maximized at the optimum number of cities. Therefore, the optimum for the residents coincides with that for a national government.

If the difference between MP (MC) and AP (AC) is caused by externalities, the Pigouvian tax/subsidy discussed by Alonso is necessary. However, our result suggests that the net Pigouvian tax/subsidy is zero at the optimum number of cities. Unfortunately, this does not imply that the optimal allocation is automatically attained by market mechanism. As seen in section 5, city sizes tend to be too big because of the difficulty in forming a coalition to create a new sufficiently large city.

Henderson (1974) formulated a more sophisticated model with three industries. The first is the export industry, which faces a fixed export price. The export industry is assumed to have increasing returns to scale. The second is the housing industry, which is assumed to have constant returns to scale. Finally, the third industry produces an intermediate good which is used as an input (called sites) to the above two industries. This industry represents the spatial aspect of cities (for example, commuting costs) which works to discourage formation of big cities. Instead of explicitly introducing spatial dimension, Henderson assumed that sites are produced with labour under decreasing returns to scale. The optimum city size balances increasing returns to scale in the export industry, and decreasing returns to scale in the site industry.

One of the most important findings by Henderson is that a market economy tends to overshoot the optimum city size because of difficulty of forming a coalition to create a big city. Our argument in section 5 is based on his observation.

Henderson (1977) extended this analysis to a spatial model and obtained (independently of our work) results similar to ours in the Marshallian externality case.

One of the major differences is that he worked with special functional forms of production functions and utility functions, whereas we assume general functional forms.

Henderson (1974) and Tolley (1974) analyzed the size of a city, considering the rest of the economy as given. Both focused on the effect of pollution taxation on the city size. Henderson showed that, since pollution taxation increases the welfare of city residents, the city size rises with immigration from the rest of the economy. In Tolley's model pollution taxation increases the city size if the externalities originate in production of nontraded goods, but might reduce the city size if the externalities originate in export production.

Serck-Hanssen, in a pioneering but little known work (1969), first obtained the condition for the optimal number of cities discussed in section 3. Adopting a framework due to Losch, he considered firms supplying their products in a space over which consumers are homogeneously distributed. Instead of assuming commuting costs, he assumed positive transportation costs for products. His optimality condition is essentially the same as ours, although in his model there is a complication arising from the fact that the optimal market areas are not circular but hexagonal in a two-dimensional space.

Mirrlees (1974), Dixit (1973), and Starret (1974) derived conditions for optimal city size in models of closed economies similar to ours. All of them assumed increasing returns to scale in the urban industry, and obtained results equivalent to that in section 3: the excess of marginal over average productivity equals the average differential rent (minus a correction if environmental externalities are present as in Mirrlees' model). Concentrating on optimal allocations, they did not analyze how the market city size is determined.

Vickrey (1977), in a very simple model, derived the result that the aggregate land rent equals the loss of a firm at the optimum, and argued that competition among cities leads to an efficient allocation. Although his analysis is not rigorous, it has the same spirit as our analysis of a system of cities formed by land-developer firms.

Arnott and Stiglitz (1975) introduced a public good which is local to a city while assuming constant returns in the production sector. In this case the optimum city size is characterized by the condition, à la Henry George, that the cost of the public good is equal to the total differential rent of a city. They also derived the following interesting formula: if the commuting costs are given by a linear function of distance (in our notation $t(x) = tx$), the aggregate differential rent equals the aggregate transportation costs in a linear city ($q(x) = q$), or one half of the aggregate transportation costs in a circular city ($q(x) = 2px$). Arnott (1979) generalized these results to include congestion in transportation, economies of scale in production, and other matters.

The central place theory originating from the seminal work of Christaller (1966) and Lösch (1954) has a close relationship with our discussion of a system of cities in section 6. Assuming that demand is uniformly distributed over space, the theory considers the spatial pattern of suppliers of goods. A hierarchical structure of central places is derived by superimposing the geographical networks of market areas for goods with different market sizes. As pointed out by Eaton and Lipsey (1979) among others, the economic foundations of the theory are incomplete in an important respect. The crucial assumption to obtain a hierarchical structure is that the location of a firm producing a good with a large market area attracts producers of other goods with smaller market areas. Under this assumption, there is a hierarchy of central places: the biggest having producers of all goods, the second biggest having producers of goods with smaller market areas, and so on. However, there is no explicit analysis of the force that causes producers to group together in this way. Eaton and Lipsey built a model in which multipurpose shopping offers an incentive for the formation of central places. Our discussion in section 6 attempts to indicate how a theory of central places might be based on the economic forces causing the agglomeration of different industries.

REFERENCES

Alonso, W., (1971), "The Economics of Urban Size", *Papers of Regional Science Association* 26, 67-83.

Aoki, M., (1971), "Marshallian External Economies and Optimal Tax Subsidy Structure", *Econometrica* 39, 35-54.

Arnott, R., (1979), "Optimal City Size in a Spatial Economy", *Journal of Urban Economics* 6, 65-89.

Arnott, R. and J.E. Stiglitz, (1975), "Aggregate Land Rents, Aggregate Transport Costs and Expenditure on Public Goods", Discussion Paper #192, Institute for Economic Research, Queen's University.

Borukhov, E., (1975), "Optimality in City Size and System of Cities: A Comment", *Urban Studies* 12, 325-328.

Chipman, J.S., (1970), "External Economies of Scale and Competitive Equilibrium", *Quarterly Journal of Economics* 86, 347-385.

Christaller, W., (1966), *Central Places in Southern Germany*, (Prentice Hall, New Jersey).

Dixit, A., (1973), "The Optimum Factory Town", *The Bell Journal of Economics and Management Science* 4, 637-651.

Eaton, B.C. and R.G. Lipsey, (1979), "Microeconomic Foundations of Central Place Theory", Discussion Paper No. 327, Institute for Economic Research, Queen's University.

Henderson, J.V., (1974), "The Sizes and Types of Cities", *American Economic Review* 64, 640-657.

Henderson, J.V., (1974), "Optimum City Size: The External Diseconomy Question", *Journal of Political Economy* 82, 373-388.

Henderson, J.V., (1977), *Economic Theory and the Cities*, (Academic Press, New York).

Livesey, D.A., (1973), "Optimum City Size: A Minimum Congestion Cost Approach", *Journal of Economic Theory* 6, 144-161.

Lösch, A., (1954), *The Economics of Location*, (Yale University Press, New Haven and London).

Mirrlees, J.A., (1972), "The Optimum Town", *Swedish Journal of Economics* 74, 114-135.

Richardson, H.W., (1972), "Optimality in City Size, System of Cities and Urban Policy: A Skeptic's View", *Urban Studies* 10, 29-48.

Serck-Hanssen, J., (1969), "The Optimal Number of Factories in a Spatial Market", in: H. Bos (ed.). *Towards Balanced International Growth*, (North-Holland, Amsterdam).

Starrett, D.A., (1974), "Principles of Optimal Location in a Large Homogeneous Area", *Journal of Economic Theory* 9, 418-448.

Tolley, G.S., (1974), "The Welfare Economics of City Bigness", *Journal of Urban Economics* 1, 324-345.

Vickrey, W., (1977), "The City as a Firm", in: M.S. Feldstein and P.P. Inman (eds.). *The Economics of Public Services*, (Macmillan, London).