Cities, Scale Economies, Local Goods and Local Governments

Oded Hochman

Urban Stud 1990 27: 45
DOI: 10.1080/00420989020080031

Urban Studies, Vol. 27, No. 1, 1990 45–66

Cities, Scale Economies, Local Goods and Local Governments

Oded Hochman

[Paper first received, September 1988; in final form, April 1989]

Summary. This study utilises a precise diagrammatic exposition to analyse issues involving the existence and stability of cities, land rents, and private and public local finance. In spite of the limitations of the exposition (such as having only two dimensions) the model is very general, primarily due to the use of cost curves in the analysis. The main new results of this study are: (1) A generalisation of the Henry George law. In its extended form, this law expresses the sufficiency of land rents for the financing of efficient local government investments in local public goods (pure or not), Piguvian subsidies and taxes, and in the internalisation of external economies of scale (except for the case of natural monopoly cities). For this study, a further extension has been made to apply to an economy with many different population groups, industries, cities, and production factors. (2) That when the basic industry of a city has scale economies internal to the individual firm, the optimal local government policy is the subsidisation of a single firm in the industry and deterrence from entry of all other firms of the same industry. In this case and, unlike that of the natural monopoly city, there may exist many such cities. (3) The identification and characterisation of internal and external scale economies in real life, and the type of city evolving from each and actual actions and procedures used by local governments which can be (and often are) used to increase efficiency. (4) The study of a natural monopoly city and the actions to be taken by both federal and local governments that are needed to achieve efficiency. (5) The existence of a criterion for the (Pareto) desirability of marginal local government investments. The criterion is based on changes in net land values. (6) The possible existence of local optimum solutions to the problem of an efficient city, which indicate the possible efficiency of major changes in the use of land in an already existing city. When the investment in such a major project is broken into stages, the initial steps in the series of investments may appear to be inefficient according to the above criterion (5), although the investment as a whole is altogether desirable.

Introduction

The reason for the creation and operation of cities has been the subject of economic research for quite some time, especially in the last two decades. In the literature we find two major reasons for the existence of entities such as cities; scale economies in production (SE) and local public goods in consumption (LPGs). Scale economies in production are effective when the industry is agglomerated in a single location, other-

Professor Oded Hochman is at the Department of Economics and The Monaster Economic Research Center, Ben Gurion University, Beer-Sheva 84105, Israel. *The author is indebted to Richard Arnott for useful comments.*

wise advantage is dissipated on transportation costs. We find this idea in Mills (1967), Dixit (1973), Tolley (1974), and others. The same effect applies to the consumption of local public good as argued by Tiebout (1956), Buchanan and Goetz (1972), Stiglitz (1977), Arnott and Stiglitz (1979), and others.

The concept of the city as a result of interaction between economies of scale in the basic industry and diseconomies of scale in the housing industry was first introduced by Henderson (1974). However, he ignored the efficiency and existence issues involved and the normative role of the local government in the attainment of those goals. Vickery (1977) in his diagrammatic model was the first to investigate these issues.

Vickery's model incorporates the essential ingredients which lead to the creation of cities, namely interaction between activities with scale economies (the export industry) and diseconomies of scale (transportation). Vickery's model is straightforward and has an intuitive appeal but its lack of sophistication limits its usefulness in analysing more complex issues. For example, Vickery does not distinguish between labour and other production factors and therefore does not recognise the double role of households as consumers and suppliers of labour, which Henderson before him showed to be a major factor in the creation of cities.

Hochman (1981) constructed a model of an economy with two types of cities, each specialising in the production of a single traded good, the production of which is subject to external SE. The model thoroughly investigated the role of local governments. The main conclusion there was that the local government has to subsidise the industry, and to finance it by local land taxes on industrial and residential zones.

As a further extension of that model, the present paper investigates two types of SE. One type are SE internal to the individual firm, e.g. the SE associated with the assembly line. We show that in this case if the SE are not exhausted entirely by the firms in the city, only one firm producing the export good will exist in each efficient city. If the city is not entirely owned by a single developer, the city manager has to intervene (to achieve efficiency) in the market operation by prohibiting competing firms from entering the city and by subsidising the remaining single firm. Otherwise the city may not be able to exist at all, and if it does, it will be too small, and produce the export good at higher than the necessary minimum costs. The subsidies should be financed by efficient (residential and industrial) land taxes which do not carry deadweight losses. Although the outcome is that each firm of the industry is in a city by itself and has a downward sloping average cost curve, the industry is nevertheless competitive and may include many cities/firms.

The second type are SE external to the individual firm but internal to the industry as a whole. One such type of SE has been investigated in Hochman (1981); Henderson (1974) and Vickery (1977) are other examples. Although we find discussions of external SE in the literature, we find scarcely anything about examples of such SE in practice. In this paper we discuss the real life nature of such SE. When the size of the industry is larger, its suppliers face less uncertainty and find specialisation less costly. This in turn leads to more varied and cheaper supplies and services to the industry. Thus, being in an industry of large scale gives an advantage to each one of the industry's individual firms. However, the individual firm, being small, does not by its actions affect those advantages. Thus the SE are external to the individual firm. Unlike the case of internal SE, when SE are external many competing firms may coexist in the efficient city, each producing the export good and facing a fixed product price as well as an upward sloping average cost curve. As in the case of internal SE, without the intervention of the city's authorities in the form of subsidies to the producers of the export goods, the city may not exist at all, and if it does, it is smaller and operating at average costs higher than it should. In this case too, subsidies should be financed by efficient local land (residential and industrial) taxes. We show that optimal local government intervention is possible in both types of SE, only if the city has jurisdiction over most of the residential areas of its labour force and most of its industrial zones.

We argue that in most current real world cases internal SE are already exhausted by the individual firms, and the prevalent effective type of SE among cities is that of external economies of scale. In Chinitz's (1961) pioneering work we find examples of this type of SE, although their external properties and their consequences have not been recognised by him.

We then proceed to a brief discussion of two irregular cases; one the natural monopoly city, and the other a city with multiple optima. In the case of a monopoly city, only one such city per industry should exist and federal intervention is required to attain efficiency. In the case of a city with multiple local optima solutions, it is sometimes necessary to perform major (non-marginal) changes in the city's land use, in order to move it from one local optimum to a superior one.

Subsequently, we discuss briefly local public goods (LPG). Tiebout (1956) was the first to suggest that consumers agglomerate into communities for joint consumption of an LPG. Buchanan and Goetz (1972) developed this idea into the theory of clubs, which was advanced further by Stiglitz (1977) and Arnott and Stiglitz (1979) into an economy-wide system of clubs. Hochman (1981, 1982a,b) elaborated further on the provision of LPG and clubs in cities. Here we investigate the provision of the LPG in an already existing city which produces an export good subject to scale economies. Our conclusion for the specific issues involved are an extension and generalisation of Arnott and Stiglitz (1979) and of Hochman (1981, 1982a,b). It is proved here that the funds needed to finance the optimal provision of LPG, above and beyond user charges and congestion tolls, can be obtained efficiently by the local government taxing residential land rents.

Free mobility of households and production factors in the economy imply that efficiency leads to the maximisation of total land rents of the city, after deducting from it the local government's expenditure on investments adding to efficiency. This provides an important practical guide to the normative desirability of a local government's action. A public investment is desirable if and only if it improves total land values by not less than the city's expenditure on it. Thus if all the land in the city is owned and operated by a single profit-maximising firm, the city will operate efficiently.

Finally, we dwell briefly on the long run in which cities are free to enter and exit and show that the Henry George law (see Arnott and Stiglitz, 1979) is actually part of the general efficiency law of zero profits (or production at minimum average cost level) whenever free entry of identical production units (in this case cities) is allowed. We show also that in the long run inefficient cities cannot survive without federal aid.

Last but not least it should be noted that the diagrammatical exposition used in this paper, besides being relatively simple and of intuitive appeal, also allows generalisation: the number and types of population groups, industries, production factors, etc., which can be incorporated in the model, are practically unlimited. Even land need not be the customary featureless homogeneous plane, and we may assume land with varying amenities and of different qualities. This paper therefore allows the identification of the critical conditions which lead to each of its conclusions.

In Section I SE and diseconomies and cities are presented. Section II deals with efficiency and the maximisation of city surplus. In Section III land rents and optimal local taxation are discussed and Section IV discusses SE and the role of local governments. The irregular cases are discussed in Section V and local public goods in Section VI. The long run efficiency is dealt with in Section VII.

I. Scale Economies and Diseconomies and Cities

A traded good is a good which can be moved and shipped from one city to another and therefore can be produced in one city and consumed in another. When scale economies are associated with the production of such a good it is termed a basic good, or an export good, since it may serve as an economic basis to a city which produces it for export, to other locations in or out of the economy.

Scale economies in production can usually be utilised through agglomeration in a single location. The assembly line is the best example of how this process works. The principle behind the assembly line is a division of the production process into many stages, each consisting of a very simple routine to perform. Each labourer specialises in the performance of one simple task which is his whole occupation.

This makes every individual labourer very skilled, efficient, and therefore highly productive in his simple task. Thus we have a relatively large group of labourers all working in proximity to each other producing more units of the final product than if each one of them was doing all stages of the process himself. However, if the labourers were located in places distant from one another, the advantage of the division of the production process into many simple routines would be overwhelmed by the increased transportation costs of intermediate products between the different locations of production.

Moreover, the advantage of the assembly line can be realised only if we produce large quantities of the good, thus avoiding partial unemployment of the specialised labour. Thus, economies of scale are fully exploited only when a large-scale production takes place in a single location.

The introduction of machines increases the economies of scale; on the one hand it involves costly and irreversible specialised investment, on the other hand it enables the production of large quantities of the good in short periods of time at the production site.

Let a basic industry include by definition all the firms in a city participating in either the production of a basic export good or in aiding and servicing such firms. Consequently different export goods, the production of which is benefited from the joint location of such firms in the same city, are included in the same industry. If a firm produces an output used as input by a firm producing the export good, both firms are included in our basic industry. This definition of basic industry is thus somewhat broader then what is usually meant by an industry. We say that two industries are unrelated if neither benefits from being located in the same city as the other, i.e. the output of one industry depends only on the inputs it employs and is not affected if the output or inputs of the other industry are varied.

When dealing in an assembly line production process, the economies of scale involved are readily recognised and the management of the individual firm is fully aware of them and takes account of them. A second type of scale economies, external to the individual firms, but internal to the industry as a whole, exists as well. The external scale economies are essentially due to better and cheaper production and marketing services and supplies, with greater diversity and lower costs of intermediate goods and raw materials, becoming available to an individual firm whenever the scale of operations of the industry as a whole is increased.

Take for instance the service of machine maintenance and repairs to the producing industry. When the industry is large many technicians (and their tools) will be needed to provide the industry with the amount of service it requires. This large scale of service demanded enables the specialisation of the technicians and their tools in particular types of service, thus improving the overall performance of the service. Additionally, a large variety of spare parts will be held in stock close by due to the quick turnover in a large industry, and the repair time will therefore be cut short, etc.

Another example of external scale economies are supplies of raw materials and intermediate goods. When the industry is large, wholesalers hold larger stocks of supplies due to faster turnover. This allows them to diversify and increase the variety of their stocks with practically no additional costs. An industry to which the diversity and speed of supply of materials is essential will therefore benefit from agglomeration in a large city. This is, for example, the situation in the fashion industry and investment markets of New York (see Chinitz, 1961), and in the fast growing electronic industry in Silicon Valley.

The introduction of standardised intermediate products, each being utilised by a wide variety of production processes, increases the advantage in agglomeration of a variety of production processes, which use the same intermediate products. This also increases the possible combinations of different production processes which benefit from locating together in a city. This in turn expands the number of possible basic industries in the economy.

Both types of scale economies alone may have led to agglomeration of all the industrial activities in one location. However, the need to provide shelter, food and services to the labour employed prevented the creation of a single, huge production site and led to the creation of many cities in each economy.

Besides basic industries which produce goods which can be shipped and sold outside the city, the city also has to provide goods and services to its labour force which cannot be traded and shipped from one city to another. Housing is the major component of such private local goods and services produced in the city for the local population which are generally provided by private markets. The production of housing is subject to diminishing returns to scale. If trips to work, shopping, and recreation made by residents of a house are included as inputs in the production of housing services, then the marginal and average costs of production of housing services increase with city size. The reason is that an additional unit of housing services in the city can be added either at the outskirts of the city where the required commuting and shopping trips to the center are costlier than elsewhere, or by raising the height of a building constructed in the midst of the city. (The average costs of construction per housing unit are higher the taller the building in which they are located.)

Therefore, the cost of accommodating an additional household in the city, while maintaining its utility fixed at an economy-wide level, increases with city size. Accordingly, while an increase in the size of the basic industry reduces the marginal inputs per unit of basic output, the costs of the marginal labour increase at an increasing rate due to the diseconomies of scale in the housing industry, causing an increase in the marginal cost of the basic good. Thus while the scale economies tend to reduce the average cost of the basic good as output and city size increase, the diseconomies of scale in the housing industry, through their effect on the cost of labour, tend to increase it. As shown below this results in cities of finite size, each specialising in a single basic industry.

In Figure 1a, the marginal cost (MC) curves of the basic (=export good) industry of a city are drawn. The variable X designates the quantity of export good produced, $MC_{Wi}(X)(i=0, 1, 2, \ldots)$ are the MC curves in which the wage rates paid to different population groups are kept constant at a rate given by the vector $\mathbf{W}_i = W_{ij}$, the index j applying to the different popu-

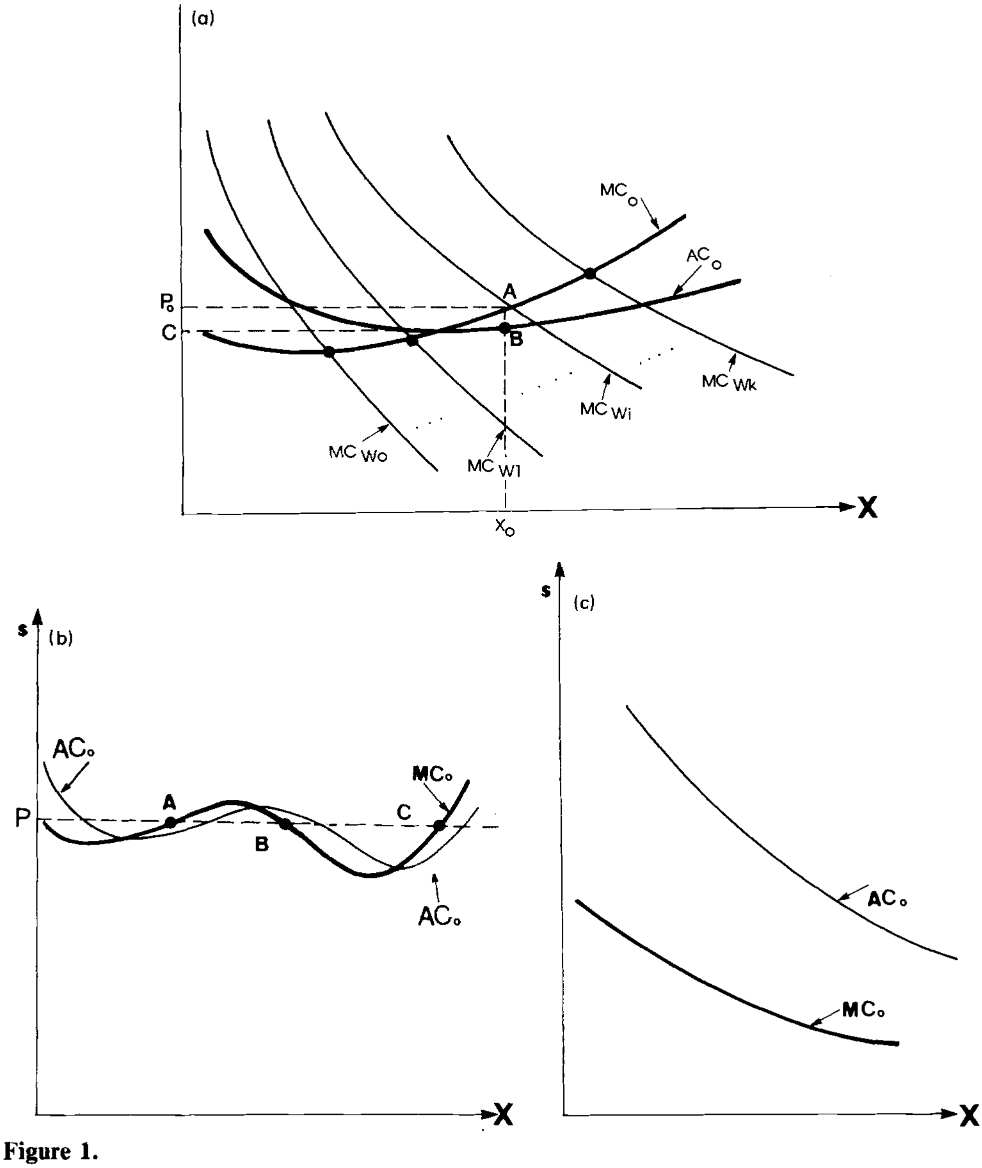

Figure 1.

lation groups in the economy.[1] The vectors $\mathbf{W}_i$ fulfil $\mathbf{W}_0<\mathbf{W}_1<\mathbf{W}_2\ldots$ The vector of wages $\mathbf{W}_0$ refers to the wages which have to be paid to the different population groups employed by the industry to enable them to maintain their respective exogenously determined utility levels, provided the city size is zero,[2] and therefore the cost of housing is the lowest possible. The underlying assumption is that free migration of population between cities leads to an economy-wide utility level for each population group. (If the utility level is not equal everywhere migration between locations of unequal utility will occur until all utility levels equalise.) Each of the $\mathbf{W}_i$ is a wage vector which would allow the labour force to maintain its predetermined economy-wide utility level in a city of size $\mathbf{N}_i$. If $\mathbf{N}_i>\mathbf{N}_j$,[3] the previous argument implies that $\mathbf{W}_i>\mathbf{W}_j$. Thus $MC_{Wi}(X)>MC_{Wj}(X)$ since the price of all factors but labor remained unchanged. The cost figures include national market prices of all mobile production factors, intermediate goods and raw materials used in the production and investment process. The cost of land, however, is the highest alternative rent from a nonurban land use. It should be noted that this alternative cost is fixed and entirely independent of the activities taking place in the city.

All the cost curves are drawn sloping downward implying that with constant wages we have economies of scale in the production of X for all relevant ranges of X. In the neighbourhood of zero quantity of X, $\mathbf{W}_0$ are sufficient wages, since to produce a small quantity of X only a small number of employees are needed and they can locate at a zero size city. As basic output (of X) increases, the number of employees and therefore of city residents increases[4] and $\mathbf{W}_0$ is no longer sufficient to purchase the additional more expensive housing and other goods needed to provide the additional labourers with their predetermined economy-wide utility level. Therefore the industry has to pay higher wages to the marginal labour, say $\mathbf{W}_i$, and MC_{Wi} is now the appropriate marginal cost function. The process of increases in wages as output increases continues, and the relevant MC curve continues to climb with $\mathbf{W}$.

The curve MC_0, which intersects each of the MC_W curves at the point at which they indicate the correct marginal cost, takes into account the fact that as the output of the basic industry increases, the wages to the marginal worker must increase as well, in order to compensate for increased marginal housing costs. Thus when output increases the added marginal costs are given by an MC_W curve with a higher wage rate. The curve MC_0 is therefore the marginal cost curve facing the city as a whole, in which both the economies of scale of the basic industry and the diseconomies of scale of the housing industry are reflected. While the utility level is fixed all along MC_0, wage rates to marginal workers are constantly increasing with output. The curve MC_0 is less steep than the MC_W curve intersecting it at the same output level, but beside that restriction, it can have any form as long as it is a single valued function. If the diseconomies associated with the housing industry dominate over the economies of scale in the basic industry then the MC_0 curve will be upward sloping, and vice versa if the economies of scale are stronger. Generally, the MC_0 curve may be downward sloping for all values of X, as in Figure 1b, or fluctuating as in Figure 1c. Nevertheless, at low city output levels MC_0 is likely to be downward sloping, reflecting the fact that in small cities the diseconomies of scale in the housing industry are negligible. In sufficiently large cities the diseconomies of scale of the housing industry are likely to be dominant, therefore at sufficiently large output levels, MC_0 is likely to be upward sloping. Between those two extremes, MC_0 may assume a downward or upward slope as depicted in Figures 1a–c. Let $AC_0(X)$ represent the total costs associated with MC_0 (total area below MC_0 up to the output level in question) divided by the output level X. This total cost is the minimum required to produce the given output level and maintain the utility level of the population constant.[5] Thus $AC_0(X)$ is the average cost curve associated with MC_0. The relationships between MC_0 and AC_0 are depicted in Figures 1a–c. The basic shape of an AC_0 curve is a U shape, but the other depicted configurations are possible as well.

We can now prove the following proposition.

Proposition 1.1. Efficiency implies that two unrelated basic industries do not locate in the same city.

Proof. Assume, as a working hypothesis, two unrelated basic industries located in a single city. Being unrelated, neither derives any benefit from the other being located in the same city, and the amount of inputs each industry requires for the production of a given output depend only on its own activity. Let N_i be the size of the labour force employed by industry i when producing X_i where i (=1, 2,...) is an index designating these two basic industries in the city. The wage rate in the city will be $W(N_1+N_2)>\max[W(N_1),\ W(N_2)]$, where $W(N_i)$ is the wage rate industry i has to pay if it were located in a city by itself. Thus the average cost of producing output X_i ($i=1$, 2) in a city of size N_1+N_2 is higher than producing the same outputs, X_i ($i=1$, 2) in two separate cities each of size N_i, i.e. $AC_0^{N1+N2}(X_i)>AC_0^{Ni}(X_i)$. This contradicts the efficiency hypothesis and is therefore ruled out, thus proving our proposition.

II. Efficiency and the Maximisation of City Surplus

Net city surplus (or gains) is defined to be the difference between the income of the city and its expenditure. The income consists of the revenue from the sales of the basic good produced in and exported out of the city plus nonearned income of residents plus net transfers (positive or negative) of federal and other public or private nonlocal sources. The expenditure consists of the value of all imports to the city of consumption goods and production factors. Also included in the cost is the alternative value of the land used by city activities, which is the highest nonurban rent which can be generated on it.

Assumption 2A (the independence assumption). Assume that all prices of imports and exports to the city are taken by the city land users as given and fixed, that utility levels are the same to all members of a population group across the economy and are independent of city of residence, and that all nonearned income and other transfers (positive and negative) are to households and independent of city of residency. In addition assume all residents of the city work in it and all workers reside in it.[6]

Lemma 2.1. Under the above independence assumption, the maximisation of net city surplus is a necessary condition for Pareto efficiency in the economy. A city fulfilling this condition is said to be locally efficient.

The rationale behind the above lemma is as follows. Since all prices of export goods are kept constant, the utility level of people outside the city is not affected by the city's activities. The utility level of people in the city is also kept constant. Therefore by maximising the surplus of the city we are creating additional income compared to the case when this surplus is not at its maximum, without reducing the utility level of anyone in the economy. This additional income can be used to increase the utility level of some people without reducing the welfare of others. When this surplus is at its maximum level in all cities in the economy and all income is distributed between agents in the economy, the economy is in a Pareto optimal state, since nobody's utility can be increased further without reducing the utility of somebody else. Thus maximisation of this surplus in any given city is a necessary condition for efficiency. A formal proof of the lemma can be found in Hochman (1981).

Proposition 2.1. Assume the AC_0 curve of a city is U-shaped, and let the city fulfil the independence assumption (Assumption 2A); then local efficiency is achieved at the output level at which MC_0 is equal to the exogenously given price of the export good.

Proof. The total surplus for the city is equal to the difference between the total revenue from sales of the export good minus the total costs required to produce this output level. When the city is a price taker, total revenue from sales of X units of the export good is P_0X where P_0 is the price of the export good. Total costs are equal to $AC_0(X)X$. The difference between these two attains its maximum value at the intersection of the price with MC_0. The rectangle P_0ABC in Figure 1a is this maximum surplus. It should be noted that the nonlocal income transfers are already included in AC_0 by affecting the level of the wages necessary to maintain the predetermined utility level of the city's labour force. QED.

III. Land Rents and Optimal Local Taxation

Assumption 3B (the full mobility and nonscarcity assumption). All economic factors except land are mobile and can be moved freely and costlessly between cities in the economy. All economic factors in a city exist in other cities as well.

This is a long run assumption. It should be noted that under it, land is the only factor whose price is not exogenously fixed and given to the city, and is determined at the local level. To see that, suppose for a moment that some production factor besides land, say labour or capital, is accumulating gains above its equilibrium rewards. This will attract more of this production factor to the city, increase its supply, and thus drive down rewards to this factor to their overall equilibrium level. Therefore, with free mobility of factors, rents can be associated only with land.

Moreover, since the supply of land of a given quality and accessibility is fixed and the only urban market for it is in the city, land can accumulate only rents. The term rent is used here in its Ricardian sense, namely the surplus left after all other economic factors are paid off. This definition of rents fits the definition of city surplus. It thus follows that under our assumptions, city surplus is part of the total land rents of the city and if local taxes did not exist, total land rents of the city would be equal to the city surplus, plus the alternative cost of the land. With possible local land taxes the net city surplus constitutes only part of total land rents, the rest are the local government's land taxes and the alternative value of land.

In this paper the focus is on actions the local government should take to achieve efficiency. We assume that local governments are the only taxing authority. The local government may tax only local factors and use its income from taxes to finance its activities.

We can summarise the discussion so far by the following proposition:

Proposition 3.1. Under the assumption of full mobility and nonscarcity of economic factors, rents are made only by land and total land rents consist of the city surplus, local land taxes, and the nonurban alternative costs of land.

Corollary 3.1. Local efficiency in a city is attained if land rents, net of taxes and the alternative cost of land, are maximised.

Corollary 3.2. A criterion to measure the desirability of a local government action. If the cost (net of associated revenues) of an action to the local government is less than the improvement in total city land rents due to it, the action is desirable, i.e. after the action the city is in a Pareto superior state compared to the state prior to the action. Corollary 3.2 thus provides us with a tool which enables us to investigate separately the desirability of different projects in the city. It is a theoretically useful result so we use it throughout this paper. This result, however, is even more important in practice.

Corollary 3.3. If the city as a whole is owned by a single profit-maximising firm, it will operate efficiently. The proof is immediate since the profits of the firm are the net surplus of the city.

Lemma 3.1. As long as taxes on land are independent of the land use and are lower than total land rents, they are efficient and carry no deadweight losses.

Proof. Ricardo was already aware of this result and therefore advocated the use of land taxes to finance government activities. The argument proving this can also be found in Mills (1972), and is as follows: the quantity of land at any given location is fixed and cannot be increased by production. As long as land owners are making positive profits and the tax is independent of the particular use the land is put to, land owners will continue putting their land to the use generating the highest net gains (the highest rent paying land use), even if they have to give up some of their rent as taxes. If, however, taxes were higher than total land rents, land owners would be losing money and therefore would rather abandon their land than keep it and suffer losses.

It can be shown that any other local tax except land taxes carry deadweight losses. Here, however, we will prove it only for a local household fixed tax (local head tax). This kind of tax is efficient as a nonlocal federal tax (lump sum tax).

Lemma 3.2. Local head taxes are distortive and carry a deadweight loss under a regime of full mobility and nonscarcity of economic factors other than land.

Proof. Figure 2 gives the income net of commuting cost of a representative household at a given location. This income is to be spent (1) on housing, *h*, which by following the accepted simplifying assumption, is produced by land only; and (2) on a composite good, *z*, the price of which is fixed and equal to 1. Let *U* be the indifference curve which fits the equilibrium utility level of this household's population group in the economy. The resident

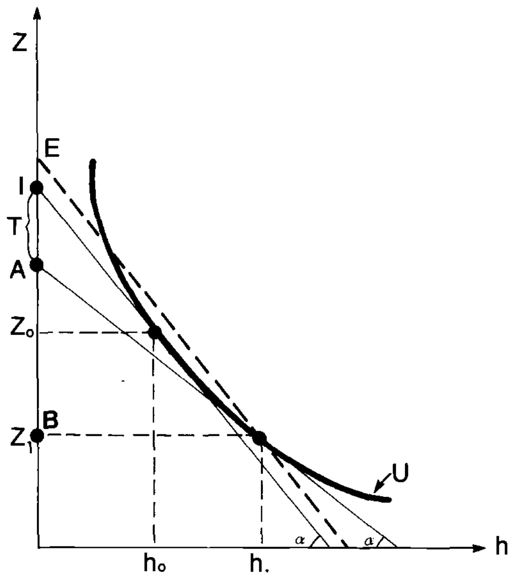

Figure 2.

household is willing to pay a rent of $R_0 = \tan \alpha_0$ (see Figure 2) per unit land (housing) and consume h_0 units of it. Suppose a tax *T* is levied on the household which amounts of IA in Figure 2. The utility level of the population group elsewhere in the economy is not affected by this tax since they are not residents of the city and therefore do not have to pay it. Consequently, neither is the utility level of the local household affected, and it is still represented by the same indifference curve *U* as before. However, since the household's money income has been reduced, it is now willing to pay for housing only $R_1 = \tan \alpha_1$, and the amount of housing now demanded is h_1 (see Figure 2). After tax, the income of the landlord from renting h_1 units of housing is AB. If no tax were levied and R_0 were the equilibrium rent, he would have obtained BE for the same amount of housing h_1. Thus the reduction in the landlord's income is equal to $BE - AB = AE$, and $AE > AI$, where AI is the government tax income from the resident occupying h_1. We have thus showed that indeed taxes on households are transferred to land rents. Moreover, the landlords' reduction in rents is higher than the government's income from taxes. The difference between the government's income and the landlords' losses equals IE, which is the deadweight loss measuring the inefficiency of this tax.

In a similar way it can be proved that sales tax on local shoppers, local income tax, etc. are inefficient as well. In the next section it is shown that tax on the export good is also inefficient. In short any tax besides taxes on land is passed on to land rents and picks up deadweight losses as shown in Lemma 3.2. Lemmas 3.1 and 3.2 and the discussion following them lead to:

Proposition 3.2. Under the assumptions of free mobility and nonscarcity of economic factors, taxes on land are the only tax base available to a local government (besides corrective Piguvian taxes) which does not carry a deadweight loss.

In the following sections, it will be shown that land rents are also sufficient for financing the local government activities intended to improve local efficiency.

IV. Economies of Scale and the Role of the Local Government

In the following proposition we discuss the normative role of the local government associated with the two types of scale economies, intended to ensure the existence and efficiency of the city. A short discussion of practical methods used by city managers for this goal follows.

Proposition 4.1. Assume the constant utility average cost curve (AC_0) is U-shaped, and the independence and free mobility with nonscarcity assumptions hold. We consider both types of economies of scale.

(a) Suppose throughout the relevant range of output that the basic industry of a city is subject to economies of scale internal to the individual firm. Then the existence and efficiency of the city are guaranteed only if either the whole city is owned by a single entrepreneur or each firm owns only its production site but the city government allows only one plant of the industry in the city, subsidises its production of basic good, and finances this subsidy by efficient land taxes. The optimal subsidy is equal to the difference between the two marginal costs of production of the basic good to the firm and to the city as a whole.

If each firm of the basic industry in the city owns only its production site, and the city exists without its government prohibiting firms from entering or without subsidising the basic industry, then the city operates inefficiently (produces the basic good at an average cost higher than the efficient one) and is too small (given that the rest of the economy is efficient).

(b) When the scale economies of the basic industry of a city are external to the individual firms, then without a local government intervention the city may not exist and, if it does, it operates inefficiently. Given that the rest of the economy is efficient, the city is then smaller than its optimal size. To achieve efficiency the local government has to subsidise the industry. The subsidy required per unit of export good is equal to the difference between the marginal costs perceived by the industry at the optimal output level and the marginal cost to the city as a whole. In this case the efficient city may include many competing firms producing the export good. No entry deterrence is required in this case.

Proof. (a). Since scale economies are internal to the individual firm, each benefits only from its own scale of operations and is not affected by the scale and number of other firms. On the other hand the diseconomies of the housing industry depend only on the total number of employees in the industry as a whole. Thus, the same labour force will produce more output if it is employed in a single rather than in several firms due to the internal SE, while

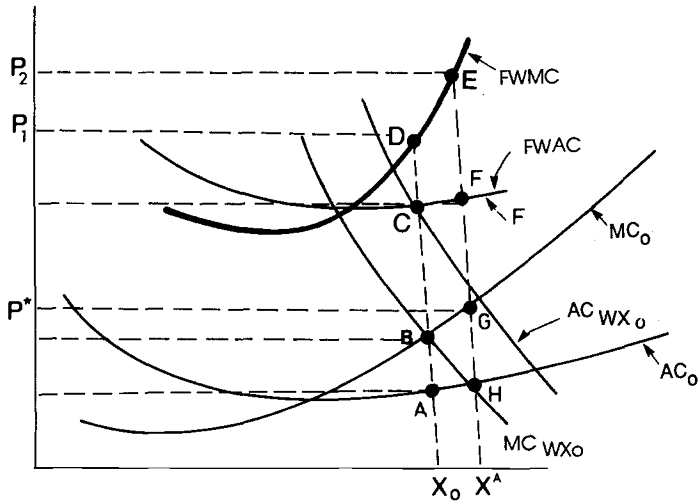

Figure 3.

the costs of its employment are the same; equivalently the same output can be produced at lower costs if it is produced by a single firm instead of by many. Therefore the optimal number of firms when scale economies are internal is just one firm per city. In this respect the situation here is similar to that of a natural monopoly, the difference being that in the current case there are many cities in the industry, each with a single firm and therefore the industry as a whole is a price taking, competitive industry.

If the single firm in the city is a wage rate taker, i.e. the firm pays all its workers the current prevailing market wages and it owns only the plant with its site and no other property in the city, then the relevant average cost it faces at any level of X is $AC_{WX}(X)$ (see Figure 3). Since the MC_W curves are downward sloping (see Figure 1) each of the AC_W curves is downward sloping as well and above its respective MC_W curve.

Each time there is an increase in the city's output, there is a shift into a higher MC_W (see Figure 1a) and the firm has to pay *all* its workers the new wage rate. Thus the average cost of producing an output of X_0, (see Figure 3) is given by point C, located on AC_{WX_0} rather than by point A on AC_0. We should remember that the wages included in the calculation of AC_0 change continuously with input, from $\mathbf{W}_0$ at zero output to $\mathbf{W}_{X0}$ at X_0, and only the marginal labour bears the higher wages. Let FWAC, the full wages average cost curve, be the locus of all points $[X, AC_{WX}(X)]$. Thus C is the only point both on FWAC as well as on AC_{WX_0}. FWAC is the actual average cost of the firm, taking into account that wages paid increase with city size. Obviously not all of that income is needed to maintain the workers' utility level fixed and only the cost of accommodating the marginal worker is that high. The costs of accommodating the rest of the labour force, living in more accessible and therefore less expensive housing (in terms of inputs not including the urban land value), are lower. The surplus income thus paid to the rest of the workers is passed to residential land owners as rents. Thus the vertical difference between FWAC and AC_0 is the average (per unit output of basic good) residential rents. The MC curve associated with FWAC is FWMC. The latter is always above MC_0 since the added cost of producing an additional unit of output includes an increase in wages to all labour already employed, while in MC_0 only the marginal labour receive the higher wages. Note that FWAC and FWMC are the cost curves of the firm producing the basic good, when this firm is subject to internal scale economies and is the only such firm in the city. In that case, FWMC is the locus of all points of equilibrium of the city since it associates with each price its corresponding equilibrium output of the city. If more than one such firm exists in the city under these conditions, then the full wages costs curve associated with the industry will be even higher than the respective cost curves associated with the single firm.

If the price of the export good of the city is below minimum FWAC, the city will not exist without a local government intervention, while an efficient solution does exist as long as the price is above minimum AC_0. Accordingly in Figure 3, P^* lies below minimum FWAC and above minimum AC_0, there is an efficient solution at X^* where P^* intersects MC_0, and no competitive market equilibrium exists. Only when the market price, say P_1 in Figure 3, intersects FWMC above minimum FWAC, an equilibrium solution without government intervention will exist. This inefficient market solution, when it exists, is always at a lower output level than the efficient solution. If the single firm has profits at this equilibrium solution, as is the case depicted in Figure 3 when the market price is P_1, other firms of the industry will have incentive for entry. New entrants will raise the full wages cost curves and by doing so increase inefficiency. Firms will continue entering and raising FWAC until P_1 intersects the FWAC curve at its minimum.

To achieve efficiency the local government has to prohibit entry to the city of more than one firm of the basic industry and thus keep FWAC at its lowest level. In addition the government has to subsidise the single remaining firm by the amount equal to the difference between FWMC(X^*) and P^* per unit output, where P^* is the given market price of the city's traded good X, and X^* is the optimal level of the city's output, attained at the intersection of P^* with MC_0. The market price plus the subsidy brings the price facing the producer to P_2, which intersects FWMC at X^* and thus leads the firm to produce X^* as desired. The average (per unit output) residential land rents in this case, are depicted in Figure 3 by HF and the industrial average rents by FE. The average (per unit output) subsidy, depicted by EG, is therefore less than HF+FE, the global average of land rents in the city, and therefore can be fully financed by efficient taxes on land rents. The remaining net city surplus, its average HG depicted in Figure 3, is then at its maximum level.

Another possibility to achieve efficiency, this time in a free market without government intervention, is when all the property in the city is owned by a single owner. Then even if the firm pays full wages to all the labour in the city, the part of it paid by the labour as rents to landlords returns to the single owner. Consequently the actual cost of labour to the firm are the wages minus rents, therefore the relevant cost curves to such a city owner are AC_0 and MC_0, and his overall net gains are equal to net city surplus. Thus the city developer, by maximising his net gains, achieve the efficient solution as well.

This completes the proof of part (a) of the proposition.

Proof (b). Now suppose that the economies of scale are external to the individual firms. The curves in Figure 4 refer to the industry as a whole. At any level of production of the industry, say X^*, each of the individual firms considers the average cost of production to be $AC_{WX^*}(X^*)$. However, not being aware of the economies of scale but only of the diseconomies of scale (e.g. limited land supply at the given site and new sites available further away from the center) the individual firms behave as if the relevant AC_w curve at X^* was $AC^e_{WX^*}$ instead of AC_{WX^*} and the associated margi-

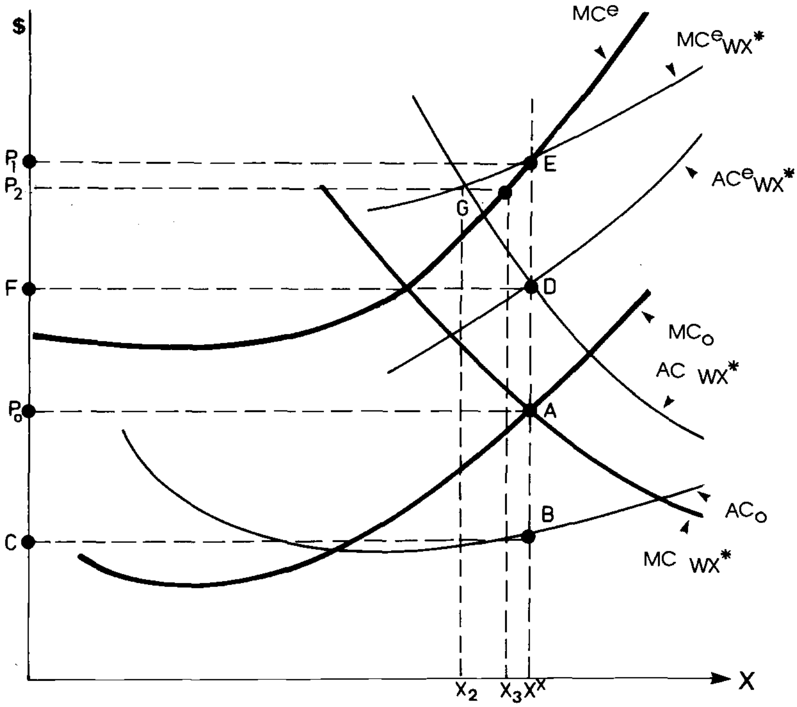

Figure 4.

nal cost curve is $MC^e_{WX^*}$ just above AC^e_W instead of MC_{WX^*} (see Figure 4). The industry then produces X^* only if the price it faces, P_1, is equal to $MC^e_{WX^*}(X^*)$, the marginal cost perceived by the individual firm at X^* (EX^* in Figure 4).

For each output level X there are different curves MC_{WX}, AC_{WX}, AC^e_{WX} and MC^e_{WX}. Let MC^e be the function which associates with each output level X the value $MC^e_{WX}(X)$ (see Figure 4). An equilibrium point is an intersection point of the price with MC^e. This is so since the relevant marginal cost according to which the individual firms make their decisions is $MC^e_{WX}(X)$ (like $MC^e_{WX^*}(X^*)$ in Figure 4). Now suppose we are at output level X^* and the price is dropped from P_1 to P_2, where P_2 intersects $MC^e_{WX^*}$ at G at output level $X_2 < X^*$. The firms will then reduce their production from the initial level of X^* to X_2. The point G in Figure 2 is, however, not an equilibrium point; at X_2 the industry will realise that the average cost is $AC_{WX2}(X_2)$ rather than $AC_{WX^*}(X_2)$, and with this realisation the firms will change their expected cost curves from $AC^e_{WX^*}$ and $MC^e_{WX^*}$ to AC^e_{WX2} and MC^e_{WX2}. The system will then move from X_2 to a higher output where $P_2 = MC^e_{WX2}$. Then expectations will change again and the system will move on. In the output–marginal cost plane only points like $(X, MC^e_X(X))$ can be equilibrium points, which means that only points on MC^e can be equilibrium points, thus when P_2 is the price, X_3 is the equilibrium output attained at the intersection of MC^e and P_2.

At the export good price, P_0, there is a locally efficient solution to the city at X^* (see Figure 4), but there is no competitive equilibrium since P_0 does not intersect MC^e. At the price P_1 there is an equilibrium point at X^*, but at that price the efficient output at the intersection of P_1 and MC_0 is much larger. Therefore if a nonintervention equilibrium solution exists, this equilibrium output level and its associated city size are always lower than the efficient output.

To induce the industry to produce efficiently the government has to subsidise it. Suppose P_0 is the market price, then X^*, at the intersection of P_0 and MC_0 is the efficient output. If the government pays a subsidy per unit of export good, depicted in Figure 4 by the segment AE, the price facing the industry will then be P_1. At this price the industry will produce X^*, the optimal output level.

Land rents in the industrial zone are then equal to net gains of the industry, their average being the different between the price, P_1, and the average cost to the industry at X^*, i.e. the average industrial rents (AIR) per unit output are: $AIR = MC^e(X^*) - AC_{WX^*}(X^*)$. Hence total industrial land rents are represented by the area P_1EDF in Figure 4. The total land rents of the city are equal to the total net gains of the city as a whole, namely the difference between P_1 and $AC_0(X^*)$ times X^*, and are represented in Figure 4 by the rectangle CBEP_1. The difference between total city land rents and industrial land rents are the residential land rents which equal, in Figure 4, the rectangle CBDF. The total subsidy required is equal to the area AEP_1P_0. Since the subsidy is only part of total land rents, the government can always tax it away from the land owners and thus cover its expenditures, leaving the maximised net city gains, which are equal in Figure 4 to the rectangle P_0ABC, to the land owners. This completes the proof of the proposition.

In our discussion, we calculated the subsidy per unit of export good, since we implicitly assumed SE generated by output. To deal with factor generated SE, we should perform in the factor market an analysis similar to the one we made in the product market in this study. Namely we should draw the upward sloping (due to the factor generated SE) curves of value of marginal product (VMP) for different fixed wage rates, calculate the fixed utility level VMP_0 curve (equivalent to MC_0), then calculate the value of average product (VAP_0) associated with the VMP_0, and proceed to finally obtain the practically same result as we already obtained in the case of output generated SE with the one difference, that the subsidy in this case should be awarded the factor which generates the SE. This result, however, also follows from the general externalistic theory which tells us that the subsidy should go to the external effect itself or its immediate cause (see Baumol and Oats, 1975). That is, output should be subsidised only if the external economies of scale are output generated (as in Hochman, 1981). In practice we know that scale economies are not always output generated and are often generated by particular inputs, like labour, an intermediate product, raw materials, local utilities and infrastructure, etc.

Consequently when the economies of scale are labour generated, labour employed by the basic industry should be subsidised; when the external advantage is due to an intermediate product, it should be subsidised; and when the advantage is in the infrastructure the municipality should provide it to the industry at reduced prices; and so on. Often the scale economies are not associated with a single generator, but with many; then, all of these generators should be subsidised.

In practice we find that local authorities do tend to subsidise factors generating scale economies; the industrial infrastructure and public services are almost always subsidised; plants which generate other additional jobs are offered attractive sites, subsidised buildings, instruction to their labour force, public health facilities and even property. (Usually deeds on undeveloped land, whose value will increase if the plant operates efficiently.)

An interesting fact worth noting is that the optimal subsidy needed to be paid to the industry is in both cases larger than total industrial land values. The differences between these two amounts are the losses of the industry which operates under economies of scale and hence with a marginal cost lower than its average cost even though the industry may not be aware of it. Consequently, in order to finance the subsidy, the local government has to levy taxes in addition to industrial zones on residential zones in which workers of this industry live. If many workers live out of the jurisdiction of their city of employment the local government may find it impossible to finance the optimal subsidy out of land taxes and it has to either reduce its subsidy payments to the export goods industry or resort to taxes other than land taxes, and hence inefficient, e.g. city income tax, sales tax, etc. which are quite common in practice. We can summarise this discussion in the following corollary.

Corollary 4.1. If a high percentage of the labour employed in the city lives out of its jurisdiction, the local government may find it impossible to finance efficiently the subsidies it has to pay to its basic industry, due to insufficient land rents as a proper tax base. In such a case the intervention of a higher authority is required to achieve efficiency. Either suburban municipalities are taxed and the funds transferred to the central city or neighbouring municipalities are united into a single, bigger municipality in which most of the local labour force also reside.

V. Irregular Cases

Consider now the case in which the AC_0 curve is downward sloping in the vicinity of the intersection point of MC_0 and the demand curve. In this case, only one city should produce the good in the economy, since this way we minimise average and total production costs of any given output. Since at the optimum output level at the intersection of MC_0 and D (see Figure 5), AC_0 is above MC_0, the city as a whole accumulates losses equal to the rectangle BCEF. In the case of external scale economies the individual producers may still be facing upward sloping, fixed wages, average and marginal cost curves, and as before an inefficient equilibrium may or may not exist. Furthermore, in an inefficient solution several cities may be producing the good rather than only one in the optimum. But even if only one city existed, the local government, unlike the case of a U-shaped AC_0 curve, could not finance from its own resources the optimal subsidy needed to induce the industry to produce the optimum output, the rectangle BHGF in Figure 5. Total land rents in the city, the rectangle CHGE in Figure 5, are less than the required subsidy, the difference being the rectangle BCEF, the losses of the city. Thus if the production of the good is desirable to the economy as a whole (a possible criterion, i.e. the total economic surplus involved is positive), a single city should be subsidised from federal resources. The federal authorities also have to regulate the price of the export product, and maintain it at the level P^*. This will also deter competing cities from entry. When the scale economies are internal to the individual firm, a similar federal intervention is required, but in this case the federal subsidy can be given directly to the firm.

Another irregularity is the case in which AC_0 and MC_0 fluctuate. In this case there may be two or more intersection points between MC_0 and a given price (see Figure 1b). Net city gains are usually maximised at one and only one of these points. Of the other such points of intersection, those in the range of an increasing AC_0 are local optima, e.g. point A in Figure 1b. The rest of the points are inefficient, e.g. point B in Figure 1b. Suppose the city is at a local optimum point (point A in Figure 1b) which is not the global optimum (point C in Figure 1b). To move from a local optimum solution to the global optimum, large-scale changes are required in the city's output level as well as size and therefore in the whole structure of land use

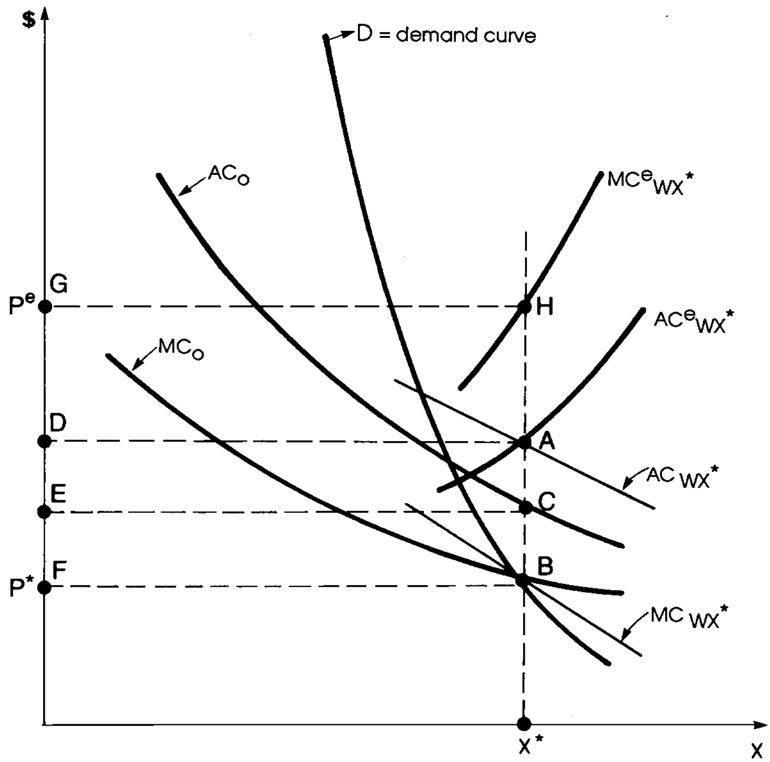

Figure 5.

in the city. The criterion of performing marginal changes in the local government policy, each change increasing the total net land rents of the city until the maximum is reached (which is useful in the regular cases; see Corollary 3.2), leads the city only towards the nearest optimum (local or global). Therefore, nonmarginal changes involving drastic changes in land uses and large public investments are required to show improvement in total land rents. When considering major changes in the city, the public investment needed is large while the exact outcome of these investments is uncertain. Therefore, the city may prefer to stay in the local optimum state unless expected gains from change are considerable.

VI. Local Public Goods

To achieve efficiency the local government has, besides internalising externalities, also to provide local public goods or local collective goods, where by local goods it is meant goods accessible mainly to the local population. In this paper we are concerned essentially with the optimal funding of these goods by the local government, and the means of determining their optimal level.

Consider a certain local public good (LPG), say public schools. To begin with, suppose the government provides a fixed quantity of schooling services independently of city size. We assume that if external effects are associated with the public good, like congestion or pollution (see Hochman, 1981, 1982a,b), the government internalises these effects by levying Pigouvian tolls on their production. Additionally, when users are free to choose whether or not to use the LPG and at what intensity, we assume that they pay users

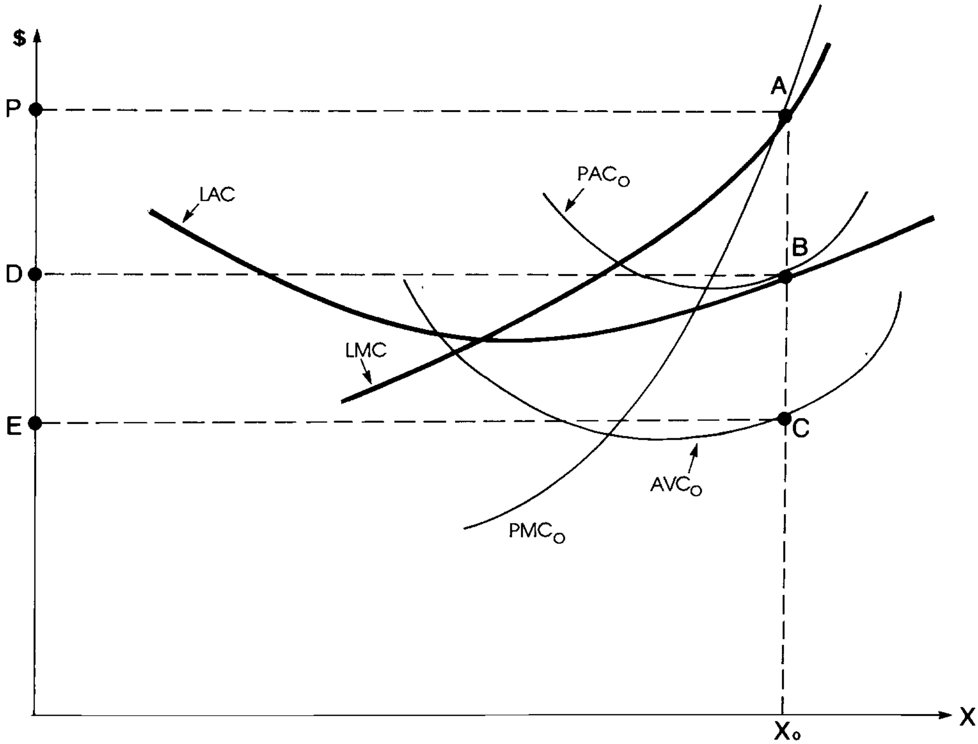

Figure 6.

charges equal to the marginal cost per user or per unit of use. Then we can derive, as in Section I, the MC_0 curve of production of the export good of the city which maintains the utility level of the city's labour force as well as the level of the public good fixed, while the wages vary with output of export good. We designate this MC_0 curve by PMC_0. Let AVC be the average cost function associated with PMC_0, which include all costs of producing the basic good but not the cost to the government of providing the given level of the LPG (see Figure 6). The curve PAC_0 in Figure 6 is the average cost curve of producing the export good which includes in it the costs to the land government of the LPG as well as the net income to the government from Pigouvian taxes and user charges related to the LPG. The differences between the two curves (PAC_0 and AVC) at each level of X is thus the average cost of the LPG to the city per unit output. It should be noted that if user charges and Pigouvian tolls completely cover the cost of the LPG, then AVC and PAC_0 coincide. In this case the LPG is a private good publicly provided. To each level of the LPG there are different PAC_0, PMC_0, and AVC curves. Our purpose is to choose $S^*(X)$, the optimal level of the LPG for each output level X. $S^*(X)$ is chosen so that the average cost of producing X will be minimised. Let the envelope curve LAC depicted in Figure 6 be the locus of all the points $[X, \min_S, PAC_0(X)]$. Thus for each X we choose the S value which minimises the value of $PAC_0(X)$. For each point on LAC there is a PAC_0 curve which touches LAC at this point. No PAC_0 curve intersects LAC, e.g. the PAC_0 curve in Figure 6 touches LAC at point B at output level X_0. The curve LMC is the marginal cost curve associated with LAC and it intersects

PMC_0 at the same output at which PAC_0 touches LAC. Thus, points A and B in Figure 6 are associated with the same output level X_0. (Through point A also passes the appropriate MC_W curve and above it AC_W, as in the previous section, although they are not shown in Figure 6.) Given that the price of the export good P_0 is fixed, the efficient output level of the city at the intersection of the price P_0 and LMC is X_0. The total net city rents (after deduction of taxes for the subsidisation of the external economies of scale and taxes to cover the costs of providing the LPG at its optimal level) are given by the area PABD (Figure 6).

Without repeating the arguments of the previous section which still hold, how land rents provide the necessary financing for the subsidisation of the industry, now we show how land rents also provide sufficient resources for the financing of the optimal level of the LPG. The average gains of the city per unit output at X_0, before the deduction of the government's expenditure on the LPG, are equal to the segment CA in Figure 6, which is the difference between the price and the AVC at X_0. These gross gains are therefore described by the area *P*ACE in Figure 6. The net costs to the city of providing the LPG are given by the area DBCE. Since total gains are equal to total residential rents (already net of subsidy to output), we can tax DBCE away from landowners and thus finance the government expenditure on the LPG without causing any distortion. This leaves net (of government expenditure on LPG and subsidies) land rents of the city equal to their maximum level as depicted by the area *P*ABD.

So far we have discussed the case in which LAC is U-shaped. As before, LAC may be downward sloping or fluctuating. The previous discussion applies to these cases as well and will not be repeated here. We can now sum up our arguments concerning an LPG in the following proposition.

Proposition 6.1. A local government can provide the (Pareto) optimal amount of a local public good (pure or not) and fully finance its operation by nondistortive taxes on land rents in addition to user charges and Pigouvian tolls.

VII. Long Run Efficiency

The intermediate run is the length of time in which the number of cities in each industry is fixed, but labour (households) and production factors, other than land, are free to move between the cities of the economy costlessly. The costlessness of mobility is due to the length of the run, in which only new investments and households are located and thus no readjustments with extra costs are necessary. The long run is the length of time in which cities can enter and exit in the economy freely and costlessly as well. The intermediate run efficiency discussed so far is necessary for efficiency in the long run as well. We now extend our discussion to the long run.

A long run supply curve of a city is its MC_0 curve beginning at the minimum AC_0 level. The industry's long run supply curve is the horizontal sum of the long run MC curves of all existing and potential cities in the industry. In that respect there is a complete parallel between our theory of the city and the standard theory of the firm, where our cities play the role of firms in the standard theory. Thus, if some cities have relative advantage over other cities due to special amenities, the long run supply curve will be upward sloping. The long run competitive and efficient solution is attained at the intersection of the long run demand and the long run supply curves. The last city to enter will be making little or no net surplus (profits) while the more advantageous cities will be making positive rents. If all cities were identical in their costs structure, then the long run supply curve will be parallel to the horizontal axis at the level of minimum AC_0. In that case all cities will be making zero surpluses. Consequently total land rents in each city would exactly equal the local government expenditure on LPG and subsidies to scale economies. When the government's expenditure is on local public goods only, the above result reduces to the Henry George law (see Arnott and Stiglitz, 1979). Thus the Henry George law is a special case of the general efficiency rule of zero profits when free entry of identical and independent economic units (in our case cities) is possible.

It should be noted that in the long run inefficient cities without relative advantage cannot survive on their own and cease to exist. In practice such cities decay slowly until, if at all, an efficient local government is elected.

Notes

1. In this paper, population groups are households having identical tastes (utility functions), the same amount of wealth (nonearned income), and the same skills. Thus, the utility level of households belonging to the same population group must be equal in the economy if free migration of population between cities prevails. In the same way the rate of return of wealth is the same everywhere since capital is mobile.
2. By city of zero size is meant that enough open space is adjacent to the plant site to locate the labour force at the cheapest density of housing. We shall designate city size by N, where the size of the city indicates both the total number of residents of the city as well as the size of the labour force. In that, we assume labour is a normal input of all cities (see note 4).
3. Note that $\mathbf{N}_i$ and $\mathbf{N}_j$ are both vectors N_{ik} and N_{jk} and that $\mathbf{N}_i > \mathbf{N}_j$ implies $N_{ik} \geq N_{jk}$ for every k, and $N_{im} > N_{jm}$ for at least one m.
4. This argument holds provided the production function is such that, to increase total output, a cost minimising industry facing constant wages will increase employment of labour. We adopt this assumption henceforth. This normality assumption holds for a wide variety of production functions. If this property does not hold the results of this paper do not necessarily hold. In practice such an industry for which the results of this paper do not apply will not locate in a city.
5. Note that this total cost curve is unusual in the sense that the cost of one of the production factors used in it (labour) is different for each unit of output. In regular cost curves, even if the price of a factor changes with output, it changes for all the units of the factor involved in producing the given output and not just for the units of the factor used in the production of the marginal unit of output.
6. Some of the following results hold only when this assumption holds. The case of intersecting cities (residents of one city working in another) can be analysed by treating it as spillover of externalities between cities.

References

ARNOTT, R.J. and STIGLITZ, J.E. (1979) Aggregate land rents, expenditure on public goods and optimal city size, *Quarterly Journal of Economics*, 93, pp. 471–500.

BAUMOL, W.J. and OATS, W.E. (1975) *The Theory of Environmental Policy*. Englewood Cliffs, NJ: Prentice Hall.

BRADFORD, D. and OATS, W.E. (1979) Suburban exploitation of central cities and government structure, in: H. HOCHMAN and G. PATTERSON (Eds) *Redistribution through Public Choice*. New York: Columbia University Press.

BUCHANAN, J.M. and GOETZ, C. (1972) Efficiency limits of fiscal mobility: an assessment of the Tiebout model, *Journal of Public Economies*, 1, pp. 25–45.

CHINITZ, B. (1961) Contrasts in agglomeration: New York and Pittsburgh, *American Economic Review*; reprined in: EDEL and S. ROTHERNBERG (Eds) (1972) *Readings in Urban Economics*. New York: Macmillan.

DIXIT, A. (1973) The optimum factory town, *Bell Journal*, 4, pp. 637–654.

HENDERSON, J.V. (1974) The sizes and types of cities, *American Economic Review*, 64, September.

HOCHMAN, O. (1981) Land rents, optimal taxation and local fiscal independence in an economy with local public goods, *Journal of Public Economics*, 15, pp. 59–85.

HOCHMAN, O. (1982a) Congestible local public goods in an urban setting, *Journal of Urban Economics*, 11, pp. 290–310.

HOCHMAN, O. (1982b) Clubs in an urban setting, *Journal of Urban Economics*, 12, pp. 85–101.

MILLS, E.S. (1967) An aggregative model of resource allocation in a metropolitan area, *American Economic Review*, 57, pp. 197–210.

MILLS, E.S. (1972) *Urban Economics*. Glenview, IL: Scott Foresman.

STIGLITZ, J.E. (1977) The theory of local public goods, in: FELDSTEIN and INMAN (Eds) *Economies of Public Services*, pp. 274–333. New York: Macmillan.

TIEBOUT, C.M. (1956) A pure theory of local expenditure, *Journal of Political Economy*, 61, pp. 416–424.

TOLLEY, G.S. (1974) The welfare economics of city bigness, *Journal of Urban Economics*, 1, pp. 321–345.

VICKERY, W. (1977) The city as a firm, in: FELDSTEIN and INMAN (Eds) *Economics of Public Services*, pp. 334–343. New York: Macmillan.