EXTERNALITY, AGGLOMERATION ECONOMIES AND CITY SIZE

David Darwent
Center for Planning and Development Research
University of California, Berkeley

Working Paper No. 109

Foreward:

Much of the material in this paper was developed in conjunction with a continuing seminar convened by the author at the Center for Planning and Development Research during Spring, 1969. I am therefore greatly indebted to the following persons whose ideas and stimulation were of great value:

Shlomo Angel	PhD. Student, Department of City and Regional Planning
Roland Artle	Professor, School of Business Administration
Douglas Lee	Assistant Professor, City and Regional Planning
Don Miller	PhD. Student and Lecturer, City and Regional Planning

David F. Darwent

TABLE OF CONTENTS

1.0 INTRODUCTION

A search of the literature reveals two main branches which have disciplinary directions, concerned with the existence of externalities, and their role in the growth and change of cities. The first stream of literature, found mainly in geography, planning, urban studies, architecture and regional science journals considers cities as generators of net economies, and discusses the "advantages" and "disadvantages," or the "costs" and "benefits" of cities in a number of fields, including municipal expenditures, the social life of individuals, costs of living, and comparative advantage in manufacturing. This stream asks such questions as what is the "optimum" size for the city? Why do some cities grow and others decline? Are there net advantages to large scale?

The second stream of literature, most of it in economic journals, considers firms as receivers (through their production or profit function) of external economies and diseconomies, and to a lesser extent as generators of the same. It asks such questions as; What are externalities? What conditions are necessary or sufficient for them to arise? How does their existence influence location, particularly in relation to urban areas? This branch of the literature is somewhat more precise and even attains rigor at times -- it is much narrower and attempts to make its hypotheses more explicit.

The existence of these two somewhat separate bodies of literature has led to some confusion in the explanation of the role of agglomeration

economies (supposedly derived from externality) in urban growth and size. It is the purpose of this paper to review some of the major contributions to these two streams of literature and for the manufacturing function to show the conditions under which externalities can be expected to arise, and their relationship to urban size. Finally, we will suggest some researchable hypotheses and ways of measuring the relative value of externalities in manufacturing in preparation for a later paper which will carry out some empirical testing.

2.0 CITY SIZE AND AGGLOMERATION ECONOMIES

Here we will evaluate some of the contributions within the "first stream" of literature identified above. Originally, the debate was largely contained in the barren discussions of the "optimum size for a city" which have dragged on at a slow pace for several decades in geographical, architectural, planning and urban studies journals. Since 1956, some of the essential futility of this discussion has been recognized, and the pet assertions of "optimum" size (which have ranged from populations of 10,000 to 1,000,000) have been replaced by a more rational discussion. However, the debate still has not led very far, and both specific hypotheses (testable) and empirical work have been rare.

2.1 Early considerations did not expressly define "optimum" and the obvious question -- "optimum in relation to what?" -- was neither asked nor answered. Later the question was asked, particularly in the infant field of urban economics developed, and the debate split itself into several sub-fields:

i) optimum size in relation to efficiency in manufacturing;

ii) optimum size in relation to social welfare and private lives of individuals in the city,

iii) optimum size in relation to municipal expenditures, and the provision of public services;

iv) optimum size in relation to development policies, and the strategy of urbanization.

In most of these fields the debate has tended to concentrate on only one side of the issue, namely the cost side, to the neglect of the benefit side. Arguments were made to demonstrate that beyond a certain population size, the costs became too great, and so cities should be limited to that size. Early writers were thus envisioning a situation as in Figure 1A. This was the case, for instance, among British town planners when arguing the "optimum" size of the new towns around London between 1944 and 1956. "Costs" were held to decrease at first with size, and then to increase, and the "optimum" size would therefore be the lowest point on the curve. Often the curve was specified a little more clearly, such as "costs of congestion," "social costs," "costs of air pollution," etc. The arguments did not, however, specify the type of costs -- total? average? marginal? -- or place a scale on the abscissi of the diagram.

Since the early 1950's, the arguments have become a little more realistic as cost-benefit techniques spread into the realm of urban affairs. Klaassen's diagram (Figure 1B) is a good summary.[1] There the other half of the argument has been added -- that agglomerations of people not only possess disadvantages -- "costs" -- but also advantages -- "benefits" or "revenues." The optimum size will be, therefore, a point where the sum of all revenues minus all costs will

be maximized. Usually this is conceived of in terms of average costs per capita of the population.

Alonso has completed this picture by adding marginal cost curves and pointing out that the important issue is to consider the differences between the sums of marginal costs and marginal revenues (Figure 1C).[2] However, this does not take us very far, of course. We must be able to disaggregate the components of the costs and revenues into meaningful categories before any empirical work can be attempted.

Another set of terms has, however, been introduced into the discussion, as a sort of substitute for costs and revenues, and these are "agglomeration (dis)economies" which have been held to be a property of all agglomerations and which have been classified into two subsets by Ohlin,[3] Hoover[4] and Isard[5] -- namely "localization economies" and "urbanization economies." These terms have had a non-specific useage and are usually invoked to cover the effect of variables whose value varies in some consistent way with urban size, for whatever function is being considered at the time.

At this point we will focus the discussion onto one of the sub-fields identified above, namely the issue of urban size, and agglomeration economies in relation to efficiency (or productivity) in manufacturing. The rest of this paper will confine itself to this issue.

2.2 Hoover's description of his three subsets of agglomeration economies is as follows:

1. Large-scale economies within a firm consequent upon the enlargement of the firm's scale of production at one point.
2. Localization economies for all firms in a single industry

at a single location consequent upon the enlargement of the total output of that industry at that location.

3. Urbanization economies for all firms in all industries at a single location, consequent upon the enlargement of the total economic size (population, income, output or wealth) of that location, for all industries taken together.[6]

A firm in a location can be expected then to have all three of these available to it (as economies or diseconomies) and the net effect will be felt by the firm. However we find this classification highly unsatisfactory, and also perhaps misleading, in several respects. Firstly, it is not clear whether these are internal or external to the firm, though we may assume that the first is internal and the second two external. Secondly, it says nothing of the conditions under which economies will or will not be available, nor does it indicate the degree of this magnitude, either relative or absolute. Finally, the classification is entirely dependent on the definition of an "industry." The S.I.C. code can be specified at 2, 3, 4, or 5 digit levels, and the degree of uniqueness of this classification will obviously affect the relative importance of the three subsets.

However, before we can proceed to define these economies more rigorously, we must look at the second stream of literature mentioned above, to gain an understanding of the nature of externalities as they affect and are affected by firms. Finally, we will return to the issue of cities and city size in relation to agglomeration.

3.0 EXTERNAL ECONOMIES AND THE FIRM

During the last 15 years, this branch of the literature has grown fast in the economic journals. Skitovsky in 1954 could write:

> "Definitions of external economies are few and unsatisfactory." (p. 143)

Now we must conclude that definitions are prolific, but in many respects not much less unsatisfactory, and Skitovsky's paper still stands as an important contribution.

3.1 A factor common to most definitions is that externalities arise when one individual's functions (firm, industry, or consumer) are affected by those of another by the provision of goods, services or disservices which are "free"; i.e., have no cost or price, and operate outside the market. In effect, if this is the sole criterion by which externality is defined, then the existence of externalities becomes a function of the degree to which accurate accounts and transactions are made for perceived benefits and costs. And further, externalities so defined could be made to disappear by government intervention -- by "arranging taxes and subsidies in order to equate private with social benefits."[8]

A second factor in the definition is that externalities arise in economic (and other) contacts between individuals when individual A's action is influenced by the actions of one or all other individuals. This is the issue of "control." Marshall's classic example of water being pumped from a coal mine fits this definition, and Skitovsky has rigorously defined it, by saying, with Meade, that externalities arise

> "whenever the output (x_1) of a firm depends not only on the factors of production (l_1, c_1 . .)utilized by this firm but also on the output (x_2) and factor utilization ($l_2 c_2$. . .) of another firm or group of firms." 9.

Thus in $x_1 = F(l_1, c_1 \ldots ; x_2 l_2 c_2 , \ldots)$, the existence of terms on the right side of the semicolon indicates available externalities.

3.2 We thus have two criteria by which externality can be defined, which can produce four possible subsets whose implications need to be investigated. These are:

A Non-market (alone)
B Non-market and outside firm's control
C Neither A nor B
D Outside firm's control (alone)

Let us consider the case in which the only definition used is "non-market" effects (A). This produces a set whose characteristics approximate Skitovsky's "direct interdependence" among either producers or consumers -- a set which in general equilibrium theory is the "villain of the piece" -- it makes a Pareto-like optimum somewhat unlikely. It includes at first glance such examples as air pollution, (residents suffer diseconomy from factories for which there is no market compensation) and the existence of a labor pool, specialized in some skills created by the existence of a number of firms, for which a new firm makes no payment. On further examination, however, the descriptive "non-market" becomes blurred, and it may be that sometimes external effects are expressed outside the market in the short run, but have a market-expression in the longer run. Thus, for example, the factories polluting Pittsburgh's air in the early Twentieth Century no doubt produced diseconomy for the residents while providing jobs and economic prosperity. More recently, Pittsburgh's difficulty in attracting firms and jobs is at least partially due to its "grimy" image which costs a great deal in direct public expenditure and in public relations. The market did in fact react (at length) with lower land values and a slower growth of the economy. This, however, is not our main point,

but needs remembering in what follows.

The subset B combines both factors -- "outside the firm's control and non-market." These externalities are a peculiarity of the production function, in which output of a firm is dependent on its own inputs and factor utilization and also on those of other firms -- but outside the market. Meade defined them, but could find "bucolic" examples only -- of a beekeeper and an apple producer for example. In fact, these "technological external economies" are rare and therefore somewhat unimportant.[10]

A third set (C) is the case in which neither of the criteria are met -- the effects are neither outside the market, nor outside the control of the firm. Clearly, this is the case of equilibrium theory and Pareto-relevant optimality, well understood, which need not concern us here.

The fourth set is more interesting and most important. It arises, as Skitovsky points out in discussion of underdevelopment problems, in which general equilibrium theory is often not relevant, Pareto optimality extremely abstract, and of little value, and in which "external economies" are supposedly very common and very important. This is the set in which market effects outside a firm's control affect its profit function, when profits depend not only on a firm's own inputs and factor utilization but on those of other firms, through the market mechanism. These are the "pecuniary external economies" of Skitovsky.[11] To facilitate comparison with the earlier definition of externality, the nature of pecuniary external economies can be expressed as follows:

$P_1 = G(x_1, l_1, c_1, \ldots; x_2 l_2 c_2 \ldots)$, where P is profit, G a constant, and x, l and c inputs.

Thus, the profits of the firm "depend not only on its own output and factor inputs but also on the output and factor inputs of other firms." Skitovsky also points out that in the context of developing nations, externalities are said to exist whenever terms to the right of the semi-colon are present.

In the remainder of this paper, we will be addressing ourselves almost exclusively to the pecuniary external economies as defined above. The most relevant question to ask is "how do pecuniary external economies arise?" -- that is, under what conditions, what criteria must be met, and how important are they? We will show that the existence of pecuniary external economies is closely related to the nature of the market in which the firm under consideration is operating, and that a particular feature of the market, namely whether returns to scale are increasing, decreasing or constant, is of great importance.

Evidently, the issue of "indivisibilities" is of importance here. From the debate between Kaldor and Chamberlain in the late 1940's, came the general agreement that the phenomenon of increasing returns (or decreasing costs) to scale in a particular factor is very much a function of indivisibilities in that factor's supply.[12] Koopmans' summary of this issue is still one of the clearest. He points out that in both the supply of labor and of capital, there are sources of indivisibility.

> "The undivided attention given by a specialized worker to a full-time task of a sufficiently challenging character produces not exactly (but presumably more than) twice as much as half-time attention (with half the training) given to the same task. . ." 13

and similarly many pieces of capital equipment have the characteristic that the ratio of inputs into their manufacture to outputs in their use cannot be reproduced at a smaller scale. There is, thus, a well-established and well understood link between indivisibility and the

phenomenon of increasing returns to scale. Koopmans also points out that the distribution of factor supplies through time and space is really another kind of indivisibility, or immobility, often not considered in "spaceless" economic models. In reality, however, when both types of indivisibility are combined, a situation arises in which the optimum strategy for investment in factor supplies at a particular place in a specific time could not be expected except by chance to approximate the Pareto-optimum of equilibrium theory in which the sum of consumers' and producers' surpluses are maximized. Hence, the existence of externalities. So the importance of indivisibilities (especially at the margin) is paramount when we are considering the

> "character of the problems posed by urban agglomerations."[14]

In effect if the economist's postulate of proportionality is held (namely that input supplies are always proportional to each other, and therefore linear, and infinitely divisible), and we assume the homogeneous flat plane of Christaller, with a ubiquitous distribution of all factor and resources, there would (says Koopmans) be no need for cities, and the well-known central place structure would not arise, even when viewed as the product of a stochastic process (as in Berry).[15]

> "this suggests that without recognizing indivisibilities -- in the human person, in residences, plants, equipment, and transportation -- urban location problems, down to those of the smallest village, cannot be understood." [16]

We must now go on to consider some of the main contributions to the literature of externality, and to develop insights into the conditions and criteria under which externalities arise between firms and industries. A later section will relate this to the earlier

discussion of urban size and growth.

3.3 Hoover addressed himself specifically to the behavior of production costs in relation to the size of the plant, and the firm, and stated three principles which have relevance for our discussions:

1. The principle of "multiples" -- due to imperfect divisibility, there is a minimum efficient capacity for any productive apparatus or unit of labor, and the smallest efficient size of a combination of factors is their least common multiple.
2. The principle of "massing of reserves." The size of stores that must be held is less than proportional to a given increase in output. The principle operates on other risks also -- as in insurance.
3. The principle of "bulk transactions" -- reduction in unit price for input supplies ordered in bulk.[17]

However, Hoover does not describe rigorously the conditions under which he expects these principles (and economics) to arise, and their effects to be felt. His classification is unsatisfactory in that it does not indicate the limits within which economies based on these principles can be expected to vary, nor does it assess their importance relatively, or absolutely.

It is evident, however, that one concept is common to all three of these principles, namely, the concept of indivisibility discussed above -- i.e., all three principles arise as a function of the imperfect divisibility of inputs to the production process, whether these inputs are labor, capital, materials, or services. In addition, the third principle is dependent on the notion of externality -- it can exist

only as a function of the conditions apparent for the input supplier, and one condition, namely that the input supplier enjoys increasing returns to scale, is necessary and limiting (though not always sufficient). If the input supplier has increasing returns to scale, then any increase in the bulk of a commodity ordered through time will result in a less than proportionate cost, and if the transport industry also has increasing returns, a less than proportionate transport cost. The resultant lower costs might or might not be passed on to the consuming firms depending upon competitive conditions in the market. It is obvious, however, that if the input supplier has either decreasing or constant returns to scale, then the principle of "bulk transactions" will not be available, and the external economy will not exist.

3.4 It is interesting to apply Hoover's principles to an industry made up of member firms, in some limited area (a region or an agglomeration), and by doing so we can illustrate further some of the conditions behind externalities. The "principle of bulk transactions" for instance, would imply that firms can gain scale economies by ordering independently of each other in bulk, and by some form of collusion, they could also order as a group, increasing the quantity of goods per order and thus gaining further economies -- this time of the pecuniary external variety. The "principle of multiples" can also be viewed in the same way. A group of firms within an industry could, for instance, employ one lawyer full time, rather than each of them employing part of a lawyer's time as needed. The indivisibility of the lawyer's services, which makes it costly to employ him on an hourly basis, would give them savings acting as a group, and again, since group action is necessary to accomplish this, we can speak of it as an externality. "The principle of massing

of reserves" can equally be applied to a group of firms. For instance, dealers in automobile parts can gain economies by having a collective warehouse, ordering in bulk, and specializing in the supply of certain parts and organizing an information system (itself a cost) between themselves. Economies would also be available directly to the consumer of the auto parts (in addition to any cost savings the firms might decide to pass on to him) if any of these economies were exploited. Specifically, the probability of a consumer getting what he wants in a single trip would rise with the number of firms and the size of the collective operation. Here then we have a complex web of inter-related externalities -- mostly pecuniary, though some not.

In all our examples, however, the common factor is the indivisibility of the input factors or goods which makes an optimal distribution of the lawyer's time improbable, and which makes it necessary to store auto parts. In addition, those economies arising from bulk transaction also imply increasing returns to scale of the input supplier. Moreover, as Koopmans pointed out, the effect of time and distance in these considerations creates farther indivisibility, or at least immobility or inertia, which makes the appearance of pecuniary external economies and diseconomies all the more likely.

We must, however, treat this subject more rigorously than either Isard, Ohlin's or Hoover's classifications will allow, consequently, they are dropped here, and a more analytical approach adopted.

4.0 EXTERNAL ECONOMIES AND THE NATURE OF COMPETITION

4.1 Assume an agglomeration in which there is an industry I consisting of i firms. Let us now examine the conditions of competition and

of returns to scale which will allow external economies to arise within the industry and between the firms in an industry and their input suppliers and output purchasers. Subsequently, we can then ask whether these conditions bear any relation to the size of the agglomeration, and from there develop hypotheses about externality and city size.

Since we are dealing here with only pecuniary external economies, we need consider only the profit function of our firms, and in particular only those features of the function on the right hand side of the semicolon in equations (1) and (3) which are outside the control of the firm under consideration. We will consider various states of competition, and also (partly by definition) various states of returns to scale -- increasing, constant, or decreasing, as in Figure 2. For simplicity, we will consider the cases in which there are only three inputs -- labor, capital and materials, and only one output -- goods.

In Figure 2 there are three possible states of competition -- perfect, monopsony and monopoly, and of course these are associated with different potential returns to scale. Perfect competition means that returns to scale are by definition constant for the firm (though this may not be the case for the industry as a whole -- a point raised later in this paper), while monopsony and monopoly imply that all three states of returns are possible, but that increasing returns are the most probable. For our purposes, the links to the right of industry I in the diagram are a reciprocal of those to the left, and will not be discussed. Moreover, certain types of competition will be excluded by definition. Thus, monopoly cannot exist in our industry I, since we are interested in firms within an industry and in monopoly there is only one firm. Secondly, a condition of monopsony on the

supplier's side (A) is of little interest, since for our purposes any condition of the supply side which can give rise to increasing returns (i.e., any non-perfect market) will be sufficient, and monopoly covers that condition.

The whole diagram, incidentally, avoids the final demand sector -- that is, the purchasers (B) are firms in an industry, and not consumers. This leaves four relationships to be investigated, which are numbered in the diagram.

Firstly let us assume a production function for our firms i, such that

$$Qi = Ai, Ki^{\alpha}, Li^{\beta}, Gi^{\gamma}$$

where K and L are capital and labor respectively, and G an indirect input of all materials and services purchased, itself of course a function of K and L, but with respect to other firms normally outside i's control -- the input suppliers. This departs somewhat from the economic convention in which a production function has only K and L on the right-hand side, but it is justifiable since we are interested in the external relationships of our firms.

The condition of perfect competition on each side of the diagram (relationship (1)) is well understood in equilibrium economics. Returns to scale are constant for the firms in industry I and also for their supplying firms. For i, the sum $\alpha + \beta + \gamma = 1$, and no individual firm has any control over another, nor can influence the market by its own actions. Returns to scale are constant, and there are no externalities available since indivisibilities in input supply are assumed away in perfect competition, and the scale of output is small. No firm has any incentive to expand, since profits are maximized at any output Q

where Q is > 0.

Relationship (2) is of more interest. Here, returns to scale in the firms of industry I are still defined to be constant, that is, $\alpha + \beta + \gamma = 1$, but the supplier of input G is now a monopolist, and so his returns to scale are increasing apparently without limit. This means that he is not passing onto his customers (Industry I) any of the cost savings being made as his scale of output increases -- otherwise $\alpha + \beta + \gamma$ for I would be >1, a paradox which defies the definition of perfect competition and constant returns. This would apply also if the input supplier was a monopsonist. Thus a condition for the continuation of constant returns in I is either (from relationship 1) constant returns to the supplier of input G, or increasing returns in G which are not passed onto firms in I. Reciprocally, when returns to capital and labor for i are constant, $\alpha + \beta = 1$, then increasing returns on input G will lend to increasing returns for the whole of I's production function, and $\alpha + \beta + \gamma => 1$. This applies to each firm within the industry and violates therefore the assumptions of perfect competition. Without collusion or control, each benefits from an economy external to itself, and the economy is entirely a function of increasing returns in the supplying industry which passes them on as reduced prices in its output. This implies that a state of perfect competition amongst firms in industry I is unlikely to be stable given a state of increasing returns of a major input-industry, G. The available externalities will be limited by the point at which returns to scale for G cease to increase, and also by the conditions of the market which will determine the degree to which G's scale economies are passed on to the firms of I.

Relationship 3 carries this point a little further, and shows that additional economies are available. Monopsony is characterized by some collusion among the firms in one industry. Returns to scale are not necessarily constant and therefore firms in the industry will be of varying sizes. The supplier is still a monopolist, whose returns to scale are by definition increasing, and so the externality due to this is also available to our firms in a monopsonic situation. In addition, however, our firms may cooperate in ordering the input G or in storing and warehousing this commodity group (however defined) and thus reducing costs,

i of the supplier in transporting the goods

ii of themselves in handling the goods.

This additional economy is available only through collusion, but is still external in that it is outside the control of any given firm, i.

Relationship 4 makes a final clarification. It answers the question can externalities exist for our firms (whether in perfect competition or monopsony) if the input supplier has only constant returns to scale and is in a state of non-collusion (perfect competition)? An increase in demand by our industry will bring forth an increase in supply by calling into existence new supplying firms, not by increasing the output of those firms, since there is no incentive to expand. Unit costs will remain the same as before, to the G-supplying industry, and to its member-firms.

The monopsonists in our industry have no source of externalities then based on increased ordering of the input G. What of the economies available to them by collusion? By getting together and ordering in

bulk with common warehousing, are economies available? Apparently not since each input supplier is operating independently and at a limited level of output. Thus the second type of economy also disappears. It would appear at this level of inquiry that a condition necessary for the availability of external economies in an industry in which returns to K and L are constant or increasing is the existence of increasing returns to scale (decreasing costs) in the supplier of inputs, the condition is not, however, sufficient in most cases.

This availability of economies is finite but depends on the opportunities for substitution amongst factors. In the case of constant returns to K and L, the existence of constant returns to the input supplier G means that neither increasing returns nor externalities are available to our firms. In the case where returns to K and L are increasing, the existence of constant returns to supplier of G means that returns to scale increase for a while, but eventually the substitution of K and/or L for G will be at a technical limit, and this is where constant returns will set in. But at no time in this latter case are pecuniary external economies available to our firms. It has been argued that the bulk/warehousing economies referred to are available even in conditions where the unit price of G remains constant. This could only be the case if the firms in the G-supplying industry are not in perfect competition -- otherwise a single supplier's limited capacity would not be sufficient to supply a bulk order, (since there would be no incentive for him to expand under constant returns). We submit that the cases where constant unit price and perfect competition are not synonymous will be rare, and so the argument is therefore invalidated.

4.2 We have discussed the phenomenon of externality so far only in

relation to the indirect inputs to the production function -- summarized by the variable G, and we find externality to be a function of the returns to scale of the input supplier, and the state of competition in the industry. We must now go on to discuss the direct inputs provided by the firm for itself -- viz., capital and labor, and the relationship between them.

4.2A The returns to scale on labor and capital are very clearly related, inasmuch as there is often a considerable degree of cross-substitution available to the firm. The employment of the two factors is thus usually sensitive to price in the factor markets. This is important in discussing the role of externalities.

A common claim by protagonists of agglomeration economies and by those needing reasons for the continued growth of cities is that the existence of a skilled pool of labor is an external economy available particularly in cities. A skilled body of labor in a city will be the joint product of the existence of all firms who need such labor up to the time of observation. The creation of skills has a clear and high cost, whether this is direct in the form of training programs, or whether the skills are acquired on the job. The existence of the skills also has a high price which will be reflected in a locally higher wage rate; the pool, or stock of such skills will persist in its location as long as employment is available and as long as returns are distributed at least equitably to labor and capital. Immobility of labor may mean that such a pool will persist even when employment is not fully available or the wage rate is depressed.

A firm which chooses to locate at the point where such a labor pool exists is therefore taking advantage of the existence of the a

pool which has been trained and assembled at no cost to itself. Its decision will presumably be based on a strategy such as this. Up to the limits of substitutability of capital for labor, the firm will have a fairly wide choice, which will be constrained by the factor prices. It might have a choice between a high investment in capital for which it would need only unskilled labor, or vice-versa. The decision in favor of the use of skilled labor and lower capital investment then results in another choice -- namely, between a location in which labor already has the skills available, at a higher wage rate, and a location in which labor has a lower wage rate but would need training in the skills needed.

Protagonists of externality would claim that given this latter choice, the greatest savings would be realized by locating at the point where labor already has the skills, and wage rates are higher. This economy results from the fact that the costs of assembly of the pool have been borne by other firms, and the new firm, at the margin thus has an economy available which is external to its own control. The economy will tend not to be reflected in the market, since one firm is gaining from the others, and no transfer payment is involved. Public policy sometimes recognizes this non-pecuniary externality as in the development districts of Great Britain, where taxes and subsidies are levied by the government in order to redistribute the costs and benefits of the training that produces skills, equitably between firms within an industry.

The notion of indivisibility is again relevant here in that the "pool" must, by definition, be of sufficient size to meet the demands of the firms in the industry if wage rates are not to rise excessively. (Such a rise, connoting a shortage of the skills would in any case be met by more training programs to create a larger pool.)

In this sense, the pool is "indivisible" or, more exactly, there is a threshold size below which it is not a "pool" for the firms involved. The well-understood factor that labor trends at best to have imperfect mobility, lends strength to this argument --- that is, there has to be an agglomeration of labor, of relevant size.

It will, therefore, be probable that such a labor pool will be located in a city rather than in a rural area, and this is one of the reasons why cities have been appealed to as generators of externalities for the firm.

> ". . .large cities with their concentrated labor market possess qualities of labor entirely absent in the country. The advantage of access to such a labor market, where any quality of labor can be found readily whenever needed, is an important element in the localization of many industries."[18]

However, while it may be true that there is a statistical relationship between the existence of a skilled labor pool and the size of a city, the more relevant functional relationship is clearly between the labor pool and the number and size of firms in the industry whose existence has made the pool possible. This is important in the formulation of hypotheses dealing with urban size as an independent variable.

4.2B Evidence of available externalities in this input to the production function is meagre. The provision of capital by banks and investors need not necessarily be an activity with increasing returns to scale. Savings on the hire of capital are more related to the creditworthiness of the borrower than on the amount he is attempting to borrow. We find the concept of externality difficult to apply in this case, and perhaps not relevant. A related issue is of importance,

however. Evidently capital is not as mobile spatially as might be at first assumed. It is certainly easier for a small firm to obtain capital in its "home" area than elsewhere, and at the same time capital tend to lack fluidity across industry borders (except at the level of the giant producers and holding companies); that is, a small firm may find it necessary to stay in its home area in order to borrow, or alternatively must attempt to borrow away from its home area only from financial sources within or connected to its own industry. We can view this type of inertia of capital supplies as a kind of indivisibility in which investors are seeking to minimize uncertainty. Under these conditions, externalities (in the form of a lower interest rate) might well be available through time or space. Both economies and diseconomies of the pecuniary variety might be evident -- however, our evidence and the resulting argument are particularly weak on this point.

5.0 OTHER SOURCES OF EXTERNALITY

5.1 An issue of some importance in this discussion revolves around the relationships between a firm and an industry, and in particular their relative returns to scale on capital and labor. In the earlier discussion when the conditions of perfect competition were assumed, a question which remained unasked was "can the industry as a whole experience increasing returns to scale while individual firms acting alone experience constant returns?" The conditions which would allow a positive answer to this question raise an important issue. The production function of the ith firm would be

$$(1) \quad Qi = Ai\ Ki^{\alpha}\ Li^{\beta} \quad \text{where } \alpha + \beta = 1$$

and that of the industry would be

$$(2) \quad QI = AI\ KI^{\alpha}\ LI^{\beta} \quad \text{where } \alpha + \beta = >1$$

K and L are capital and labor

Equation (2) is thus the summation of the individual five equations (1). None of the individual firms have any incentive to expand or to buy up other firms since profits are maximized at any level of production greater than zero. Thus, if the market demand is expanding, we would expect supply to increase by 'n' the number of firms, and not by an increase in output of an individual firm. But if the demand increases and the industry as a whole reacts by meeting this increased demand with an increased output, then it must be experiencing increasing returns (or decreasing costs) to scale; otherwise there would be no incentive to increase the supply. This implies that for the industry $\alpha + \beta$ are >1.

Evidently, this can only happen if $A(I) = f(n)$, that is, if the constant is not a constant but a variable related to the total number of firms in the industry, which increases at a rate greater than (dn). Another way of looking at this is that we are imputing another element in the production function (say C) which is itself a decreasing function of the total number of firms in the perfectly competitive industry. However, if $A(I) = f(n)$, then as a new firm joins the industry to increase the output in response to greater demands, either this ith firm produces a given output at lower cost than all other firms, or else such lower costs are spread amongst all firms, otherwise $\alpha + \beta$ for the industry could not be greater than 1. This latter alternative would seem reasonable if we allow (as suggested) the adoption of the input factor C into the production function of all the firms. Since

C (a cost) is decreasing as the number of firms increases, then each firm experiences decreasing total costs for its (constant) level of output, which is in effect a form of increasing returns to scale. In fact such a situation would provide the incentive for firms to expand, thus violating all the assumptions of perfect competition, and therefore destroying the equilibrium.

This we would suggest is an illustration of a real "agglomeration economy," and explains why perfect competition is so rarely found outside the economic journals. However, we must further illustrate this new element in the production function C. We can call it the cost of "connectivity" between firms within the industry. It decreases as 'n' rises, and this implies in spatial terms that it will decrease as the firms become more dense in a given agglomeration or region, and will disappear at the point where all firms are in the same location and when firms are economically so close as to be indistinguishable from each other. Thus, if we have eight firms in an area, called collectively an industry and if we assume that each link between them has a unit cost of 1, which can be reduced to zero by the firms sharing one location, then with eight separate locations, the sum of the cost of connectivity will be $64 - 8 = 56$, in four locations $64 - 16 = 48$, and in two locations 32, until all firms are in one location, when the cost will be $64 - 64 = 0$.[19]

We can further illustrate these conditions and criteria by looking at the relationship between the capital and labor inputs in a two-variable production function under conditions of decreasing, constant or increasing returns to scale for an industry as a whole, through the use of transformation curves.

In figure 3A, each isoquant represents a series of equally optimal points along which K can be substituted for L at a given level of output Q. The single optimum is further defined as that point (P) of substitution at which the q^{th} isoquant meets slope $-(r/w)$ where r is rent on capital, and w the wage rate. The line joining these points $P_1 \ldots P_2 \ldots P_n$, is then the familiar expansion path of the firm, which may or may not be a smooth curve. However, in this analysis, the diagram represents an industry made up of a number of firms, and the cases 3A, 3B, 3C are of increasing, constant and decreasing returns respectively. Scale economies or diseconomies exist therefore only in 3A and 3C, and then only in moving between any two isoquants as the industry grows. Figure 3A is a non-perfect market, and for any given output Q the optimum number of firms in the industry will always be 1 (monopoly), since economies are always available by increasing output. The existence of a number of firms (monopsony) would mean that economies for the industry as a whole would be fewer. For example, in our diagram (3A) at output Q = 20, one firm would absorb 4.7 units of capital and 4.3 units of labor. For the same output, two firms would each need 3.9 units of capital and 3.7 units of labor. The savings by combining the output into one firm are thus 3.1 units of capital and 2.9 units of labor. The availability of externalities is not, however, affected by the number of firms in the industry whilst returns continue to be increasing. That is, scale economies and externalities are not mutually exclusive, but the latter cannot exist without the former for the industry as a whole, though this may not be true of course for the individual firm(s). In many industries the relationship between K and L is such that at higher levels of Q the industry's (and this applies

to firms also) expansion path will tend to be optimal with relatively larger inputs of K, and smaller inputs of L -- this is in the nature of what we call "technological progress," and viewed in extremes (the contrast between very small and very large firms, or low Q and very high Q) has other connotations such as the concept of "technological maturity" and the movement of industries from city center to suburban areas which will be reviewed later.

A common claim in the literature of urban economics is that, as a firm grows in size it "internalizes" the available external economies. If this were valid, we would then expect an inverse relationship between firm size and economies, and that decisions about investment, location, and company policy will be less influenced by factors outside the firm's control. However, in some senses this is a fallacious notion. It is true that (say) as an automotive plant grows in size, it may choose to integrate vertically performing for itself those activities previously carried out by others as suppliers or buyers. Thus, at the output end, Ford may choose to control its distributors (and their prices) and at the input end, may want to buy up the pressed-steel firms which supply body shapes. However, short of integrating the whole economy under one head, the firm will still have input suppliers and perhaps output purchasers other than final demand. Given this, externalities may still be available. In effect the larger firm merely pushes backwards the point at which it must rely on other firms and industries to supply it, but the possibility of external economies evidently still remains -- what changes is the point at which we must begin to look for them. The availability of external economies may well be related to the proportion of the total

input value which the firm has control over, but this may not necessarily be a function of firm size, and will in any case vary considerably between industries.

In Figure 3B, returns to scale are constant, and there is no way of determining the optimum number of firms for a given output Q. However, this is the state of perfect competition and the assumptions and implications of that state apply. No firm has any incentive to expand, and output will be increased by adding new firms to the industry. As pointed out above, in this case external economies are not available, since their existence (as a function of increasing returns to the input-supplier) would destroy the equilibrium which is essential for perfect competition (and constant returns) to exist.

Figure 3C has a certain academic interest, but it does not provide us with much insight, since a state of decreasing returns could only lead to the bankruptcy of both firm and industry. Even if returns were at first constant or increasing and only became decreasing at a certain level of output Q, continued expansion into an area of decreasing returns would, to say the least, be irrational in a competitive market. Externalities would be available theoretically, but the firms would not experience them since they must always decrease their output (and therefore their input orders) hoping that returns will eventually be constant or start to increase. Input suppliers would experience them as declining customers, of little value.

Thus, the only situation of relevance to the availability of external economies through the production or profit function is one in which returns are increasing for the firm and/or the industry. Within this the conditional necessity is that either the input supplier

must have increasing returns, or that indivisibility in the supply of labor and/or capital is evident. Other helpful conditions have been described above in terms of the states of competition amongst firms in the industry and between them and their suppliers of direct inputs (capital and labor) and indirect inputs (materials and services).

5.2 Raymond Vernon in summarizing work on the New York Metropolitan region study hypothesizes a close tie between industries which are dependent on externality, the incidence of uncertainty in their economic lives, and city size. New York appears to be the home of these "technologically immature" industries, characterized by rapid innovation, uncertainty in demand, unstable output, a wide range of unstandardized products with a high information content, and so on. Publishing, the fashion industry, radio in the 1920's, TV in the 1940's are all examples of this kind of activity.

> "Businessmen in lines of this sort must suit their methods of operation to such uncertainties. They cannot commit themselves to specialized machinery; for specialized machinery, though well able to turn out long runs at low cost, usually is not easily adapted to swiftly changing products." 20

Consequently, such industries tend to have transformation curves heavily weighted towards labor, and they expand output not by investing in capital in order to raise labor productivity, but rather by increasing the labor force as and when necessary. This, claims Vernon, goes a long way to explain why New York has a poor record of growth of labor productivity, whilst employment has either continued to expand slowly, or is at worst static; and we might add that returns on capital -- that is the capital-output ratio, or capital productivity -- is reciprocally growing fast. Because of the restraints on capital investment (and high

interest rates are common when the borrower is faced with great uncertainty), the industries which have the above characteristics tend to consist of many small firms rather than a few large ones, and tend to use a great many independent supplies in this production function. That is, the proportion of total inputs which is out of the firm's immediate control is high, and those inputs are many, diverse, and fickle.

> "In businesses producing such things, the entry of new competition is usually easy; a new firm need not have much in the way of capital because there is no need or opportunity for heavy investment in equipment and because there are not dominant brand names whose lead must be overcome. But exit may be just as easy as entry; the future is precarious and every possible advantage in speed and flexibility must be fully exploited." 21

The turnover of firms in these industries is thus likely to be high, but as Vernon points out, the industries which have these characteristics do not grow faster in New York than elsewhere in the long run. The reason for this is in Vernon's concept of technological maturity. In the 1920's, the radio industry was concentrated in downtown New York, in lofts, and attics and other cheap rental spaces; there were many small firms and few large, and turnover was great, whilst capital investment per employee was small. By 1940-50, the techniques of producing radios had developed to the point where the market was more assured, costs were lower, and sets were being produced by the "on-line" system in large runs. This was achieved by ever greater inputs of capital relative to labor, and ever greater inputs of space per unit of output. The threshold size of firms grew, entry became more difficult as large quantities of capital were needed, but exit was not so readily assured by failure to respond quickly as it was in the earlier period.

The result in the radio industry has been that by a process of net

growth rather than relocation, the industry has shifted out of central New York to the suburbs, and out of New York altogether to smaller agglomerations. The decreased dependence on skilled labor, on services and materials supplied by others in small quantities, and on access to information about the rapidly changing market is replaced by an increased dependence on supplies of non-skilled, cheap labor, on large areas of flat land, and thus, in effect, the internal resources of the firm replaces earlier reliance on externalities. Both through time for the industry as a whole, and in the shorter run through time for the individual firm, the optimum balance of factors K and L in the production function shifts increasingly towards K as output Q moves from lower to higher value isoquants. The firm and the industry both grow through raising the productivity of labor by the application of increasing amounts of capital; and capital-output ratios can be expected to grow more slowly, or even decline.

However, if we look at one particular agglomeration or at the center of an agglomeration of large size in contrast to its suburbs, we experience the technological maturation as a series of waves diffusing from a center outwards, each replaced by a newly generated technology arising in the lofts and attics of the old one. Or at least this was the case until the special features of our new electronic technology tied to the expanding military budget itself became mature, and generated further technological advances in its own laboratories in its present locations (the west and midwest, outside the big cities in "defense industry" centers).

All this must, however, be regarded as somewhat hypothetical, since to date data and empirical evidence have not been exhaustively manipulated towards establishing the truth. Vernon is clearly implying a

link between "dependence on external economies" and "technological immaturity" which in a more specific form is a researchable issue. A further implication, however, is that there is also a direct relationship between the size of the agglomeration and the availability of externalities, and the balance of the factor inputs, K and L, in the production function.

We would expect a logical continuation of Vernon's argument to state that since the characteristics of technological immaturity include a weighting of the production function towards labor and away from capital, then we should be able to say that dependence on externalities is a function of this relationship -- such that the dependence declines as the balance of inputs shifts more towards capital through time. But the implications of our earlier discussion suggest a different hypothesis, namely that dependence on (or availability of) externalities is related to the proportion of total input value which comes indirectly to the production function -- that is, the value of goods and services supplied by others in relation to total input values. We also claimed that the availability of externalities will be related to the number and/or size of firms in an industry in the particular agglomeration under discussion. These ideas in turn suggest a strong link to growth pole and growth center ideas, with the connotation that the availability of externalities in an industry has much to do with the types of links (their value, frequency and range) between the industry and others across the input-output matrix. Analogously, the same would apply also to firms within an industry.

An issue which emerges from all this is that of the relative importance of city size and number of firms in an industry in

explaining the availability of externalities. Obviously, the two are in some way correlated, and it is an oversimplification to imply that one can measure "urbanization economies" available to a firm by the population size of the agglomeration it is located in, whilst "localization economies" will be measured only by the number of firms (and their size) in the industry in that agglomeration.

Before we go on to consider some specific hypotheses and ways of measuring the value of net externalities, one further issue must be raised. At the continuing seminar convened by the author in the CPDR (referred to above), a hypothesis was advanced by Roland Artle, from a discussion of Mill's paper[22] that the distribution of firms within an agglomeration is very much a function of their available returns to scale. Artle suggests that firms (or industries) having been borne in the central areas of large agglomerations experience increasing returns to scale during their early development. Any innovations that can be made by capital investment will, at this stage, yield much greater than unitary returns. Moving between isoquants, returns to labor increase as the balance of factor inputs begins to shift from L to K (Figure 4). At some point (p), at which Vernon would see approaching technological maturity, increases in Q absorb more factor inputs per unit and the balance of inputs is still shifting towards K. Labor productivity grows at the expense of capital productivity, and returns to scale become less dramatic. It is at this point that location theorists and ecologists appeal to the needs for larger areas of cheaper land to explain the shift of location to the suburbs. Freedom from dependence on skilled labor also releases the firm from a central location which has access to the whole labor-shed. As the firm moves through the

isoquants, returns to scale become less and less dramatic as capital becomes a substitute for labor, until the point (X) at which we can speak of the change as a difference in kind rather than degree (Figure 4). There the giant plant on the outskirts of the metropolitan area is designed to operate at a given capacity and economies of scale are not possible while operating at that capacity, but only below it. Further output requires jumping from one isoquant to the next, due to the great indivisibility of the modern automated plant. Returns are then constant at a designed level of output Q, both for K and probably for L.

6.0 SOME RESEARCHABLE HYPOTHESES

The complexity and confusion represented in the last section are indicative of the many conflicting hypotheses which it is possible to specify in order to develop the notion of agglomeration economies. Here we will classify some of the more common ideas into several sets, and then close by stating a few hypotheses to be tested and evaluated in later papers.

6.1 Evidently we are dealing with a phenomenon which is simultaneously inter-spatial, inter-temporal, and dynamic. Hypotheses can, therefore, hold constant the spatial variable to detect temporal changes, or vice versa, or they can deal with time and space simultaneously. Other distinctions to be made include subdivisions of the spatial dimension to allow specification of intra or inter-metropolitan distributions, and axiomatic variations concerning the distinction between a firm and an industry.

Slow progress in the field of measuring externalities and their affects can be attributed to several factors. Conceptual confusion has

been prevalent, data have been inadequate, and relationships highly multi-variate and complicated, such that the marginal affects we are interested in have been difficult to isolate.

The problems of measurement will be dealt with fully in a later paper, at which time the choice of measure made here will be defended and justified. However, we can say that the most relevant measure from the point of view of the firm would be one which subtracts all input costs from output prices for a year, while standardizing for variations in the wage rates, etc. This measure would have spatial and temporal variation which in regression could be related to all other hypothesized factors. We would then have a relative, not an absolute, measure, of the variations in "value added," through time or space or both, and as a function of other variables. This is not a direct measure of course, but it has the merit of being useable with the data available to us. Moreover, we have a choice of attempting to hold other factors constant (ceteris paribus) while looking for variations in "value added" or of treating all variables simultaneously in regression or some other multi-variant technique. Most work on the problem to date has, in fact, used some variation of the "value added" measure. Other approaches have relied upon the "residual" value added remaining after all the rest has been accounted for, this residual being post hoc identified as the effect of agglomeration economies. Clearly, this latter approach is much less satisfactory since it forces us to rely on the first criteria of definition put out in this paper (Section 3.2), namely that externalities are those non-accountable effects. In the paper which follows we are to concentrate only on the

evaluation of "pecuniary external economies" as defined above, and the existence of technological externalities will be assumed away.

6.2 We can classify hypotheses about externality and city size as in Figure 5, in which the sole dependent variable is the measure, "value added in manufacturing."

FIGURE 5

Variations in the Measure, "Value added in manufacturing;" Classification of hypotheses.

AGGLOMERATION \ INDUSTRY	A. Between		B. Within	
	1. At one time	2. Between times	1. At one time	2. Between times
X Between	1	2	5	5
Y Within	Artle	2,3 Vernon	4	4 Vernon

We can now look at some useable hypotheses which fit into the boxes of Figure 5, from which four are selected for testing later. In the listing of the classification, the first term will represent the object of the dependent variable (city, or industry).

1. A2X.(A1X) Industries highly localized in large agglomerations will have a correspondingly slow growth in labor productivity, and lower levels of capital investment per employee than industries which are either localized in smaller agglomerations, or not particularly localized with respect to agglomeration size. Changes in the localization of industries through time will be correlated with changes in the rates of growth of both labor productivity and capital investment.

Note: In this hypothesis, the dependent variable is not value added, but rather some measure of labor productivity say value added per employee, or value added per man hour. The hypothesis is derived from the notion (by Vernon) that at any one time, industries localized in a large agglomeration are those dependent on externalities in supplies of inputs. These industries are also the "technologically immature." This hypothesis is dependent on some effective measure of localization, of which there are a number available.

2. A2Y. (a) Variations in the distribution of industry within an agglomeration (of large size) will be such that industries in the center will have the characteristics of hypothesis 1, and industries located in the periphery will tend to have the reciprocal of these characteristics. We will expect specifically that labor productivity will be low and grow slowly along with capital investment per employee, in center-locating industries, while the opposite will tend to be true of periphery-locators.

3. A2Y. (b) A hypothesis similar to the one above for which we are indebted to Roland Artle, which states simply that within an industry, distribution between center and periphery will be such that returns to scale will be greater for center locators than periphery locators whose returns will be almost constant. We can add that in a two-variable production function, the center locators will have a transformation curve optimum loaded towards labor, and those in the periphery towards capital. In the form (A2Y), this hypothesis would assert that through time industries whose returns to scale begin to increase will tend to relocate (directly or by net growth efforts) in the periphery; at the higher Q's involved capital inputs will grow

relative to labor input.

4. B1Y B2Y. These hypotheses are identical to those in 3, except that they apply to firms within an industry.

5. XB1 (XB2) A "cross-sectional" hypothesis of variations in agglomeration economies between industries will assert something of the following. Within an industry, variations between agglomerations in value added (and in productivity), are a function of the population of the agglomeration and the number and/or size of firms in the agglomeration in the industry under consideration. In its (XA2) form, this hypothesis states that changes in the value added (and productivity) variables are a function of population change, and change in the number and size distribution of firms.

7.0 CONCLUSION

Conclusions are premature in a paper such as this, since subsequent papers will attempt empirical testing of the hypotheses. In particular, the issue of the relative importance of "urbanization" versus "localization" economies, or indeed the validity of the distinction itself, remains to be tested thoroughly. At this point, however, we would question such a distinction from the firms' point of view, though it may be of more value when comparing cities and their rates of growth and change.

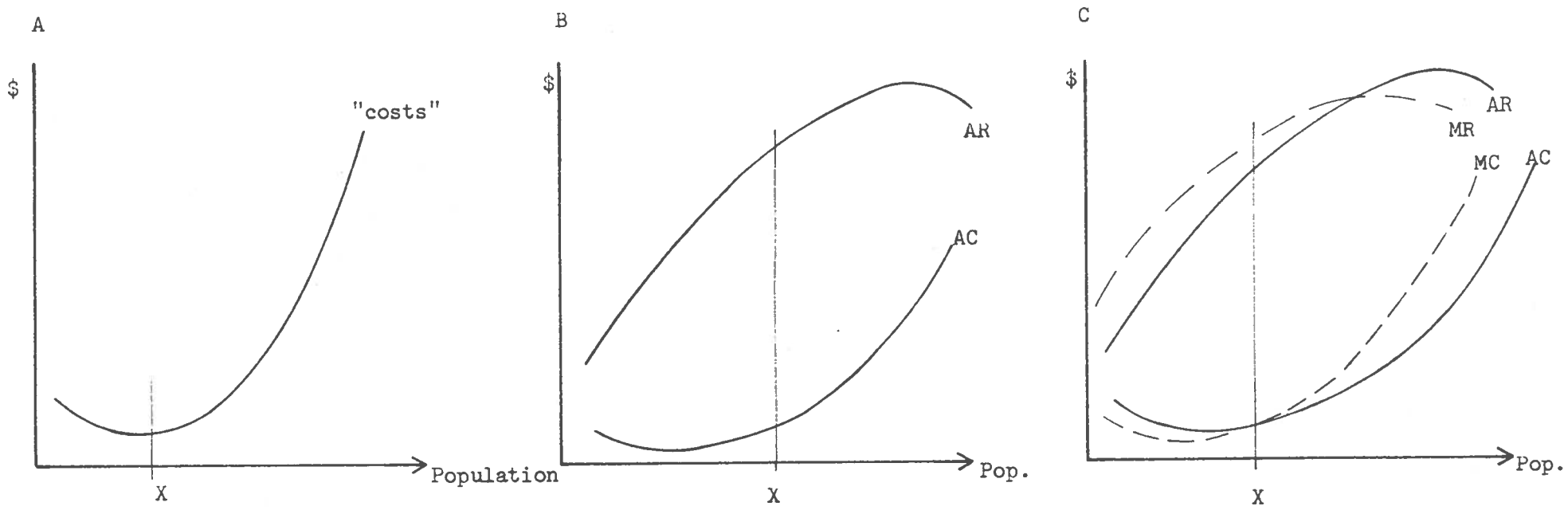
We find Skitovsky's "two concepts" to be of great value (still) and have undertaken further exploration of the "pecuniary externalities" as they affect and are affected by firms in a market which has spatial dimensions organized discretely around and within agglomerations. Clearly this does not go far enough, since the issue of the existence

of "technological externalities" and their relative value must also be raised. Equally the use of a discrete measure of spatial variation such as the agglomeration is justified only in the context of current interest in the role of large population centers. Space and time are both much more realistically dealt with as continuous variables and conclusions can be invalidated when discreteness is introduced into empirical testing.

In developing Hoover's classification of scale economies into a set of criteria by which the availability of externalities to industry can be recognized, we are attempting to develop insights into the links between returns to scale for factors (productivity), the states of competition and the availability of externalities. This argument lacks precision and rigor, and should be regarded as an early attempt at development in this field. Clearly, the need here is for some sound economic analysis and a more powerful theoretical statement. The barriers to this at present are conceptual -- it is difficult to bring into the argument spatial variables which are only inadequately understood, and to develop models which depart boldly from the confines of equilibrium theory.

Our final section which reverts to the issue of externality, or agglomeration economies, in relation to cities and city size, only begins to scrape the surface of what is evidently a very rich and interesting quarry. Progress here will, however, depend on some sharp empirical work, much of which awaits improvements in available data, or perhaps the special collection of data at a micro-scale in conditions which rule out much of the unwanted exogenous variation of the related factors. An attempt to develop this empirical part of the inquiry is described in a subsequent paper in this series.

Figure 1. Some Concepts of Costs and Benefits of City Size

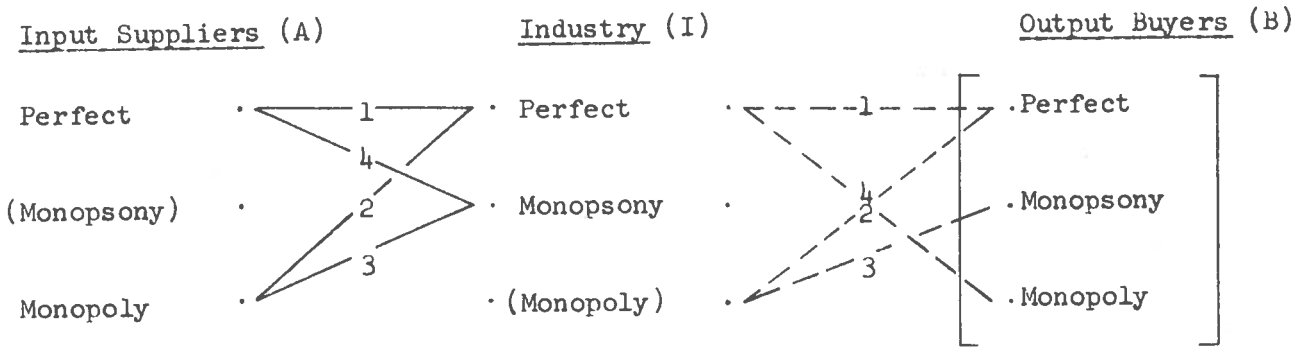


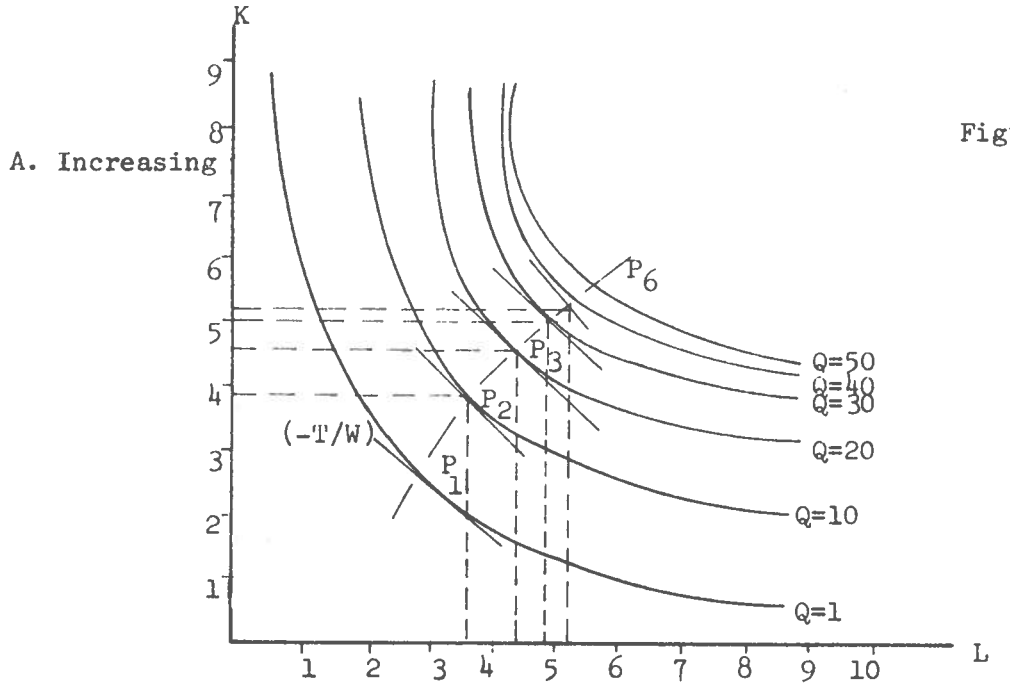
Source: Klassen op. cit.

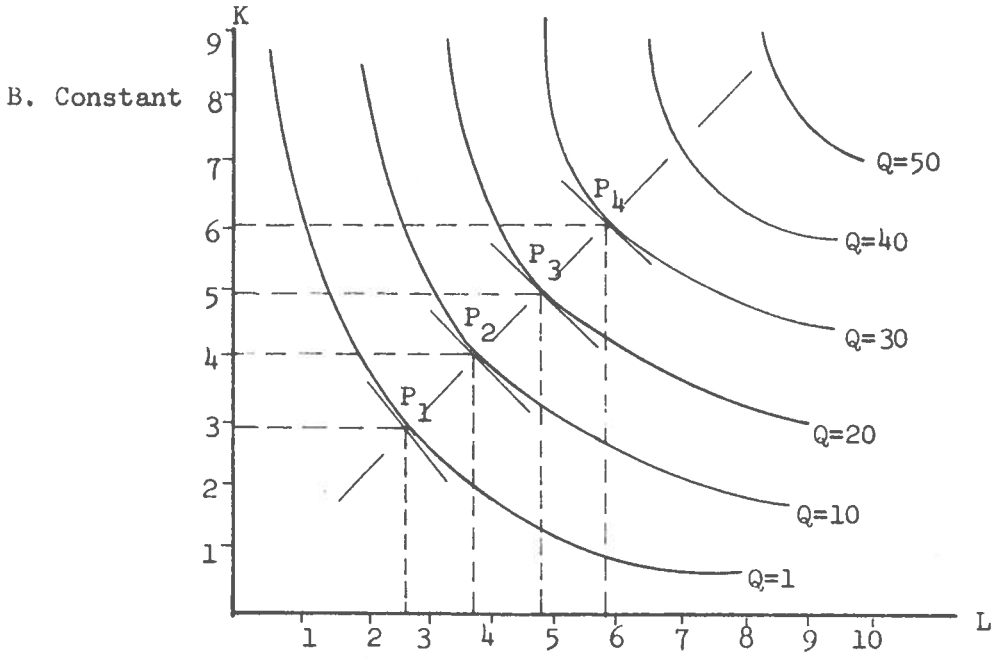
Source: Alonso, op. cit.

AR = Average Revenue
MR = Marginal Revenue
AC = Average Costs
MC = Marginal Costs
X = Optimum

Figure 2. States of Competition in Relation to Externality







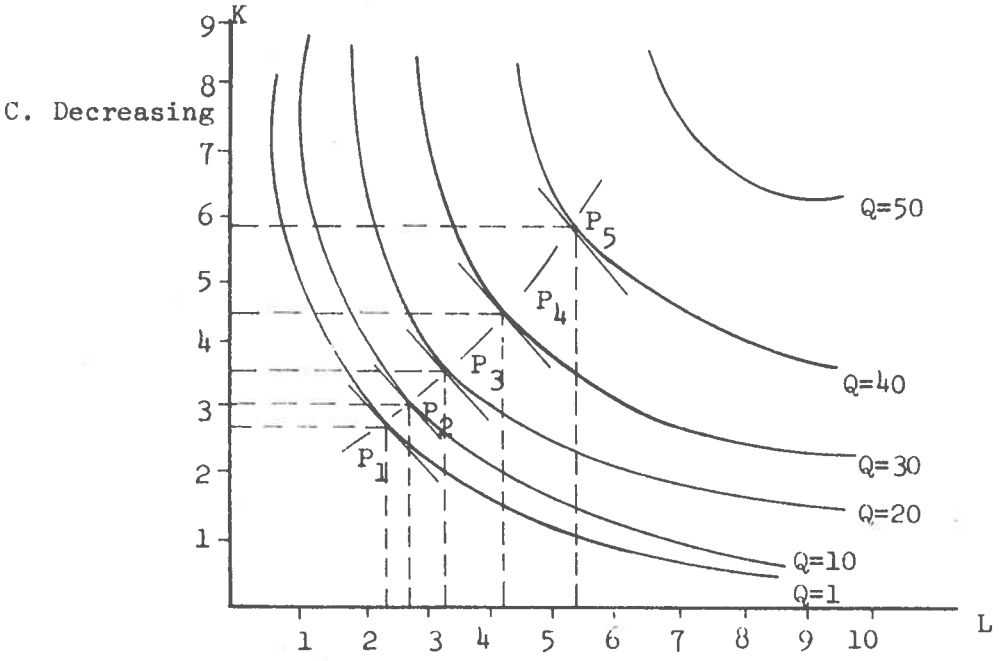


Figure 3. Industry (I) Curves o
Increasing, Constant
Decreasing Returns

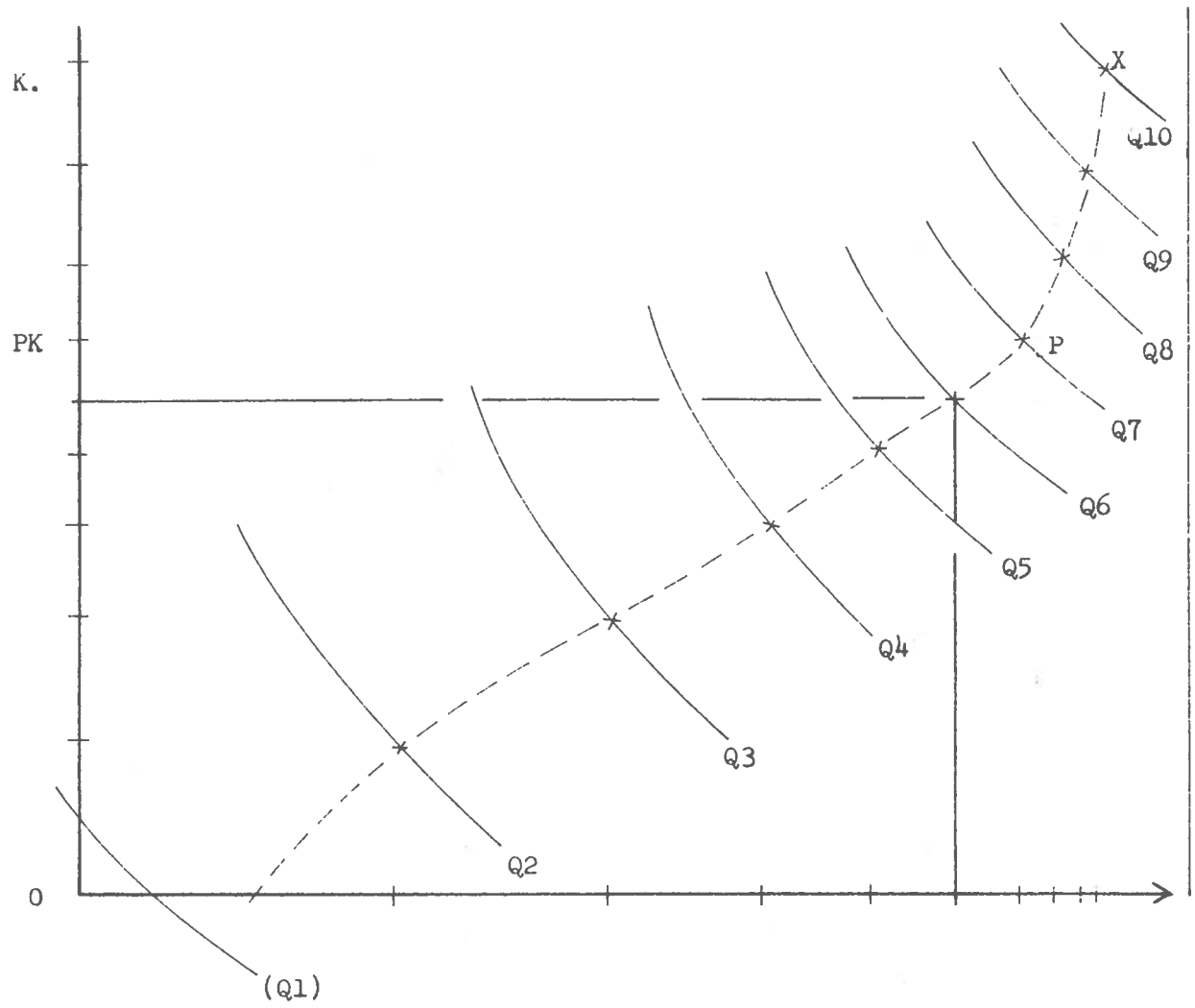


Figure 4. Hypothesis of the growth of firms within large agglomerations

NB Between O + P the sum of K + L is decreasing, after P the sum is constant. However, the expansion curve OX is such that K replaces L gradually by decreasing at a rate < L below P, and by increasing at a rate > L above P. At P, the sum of the inputs required to move from Q7 to Q8 is exactly divided between K + L. Their marginal costs are equal.

Notes and References to Text

1. Alonso, W. This point has been brought up by Professor Alonso on on several occasions in discussion at the Department of City and Regional Planning, Berkeley.
2. Klaasen, L. H., "Regional Policy in the Benelux Countries," in Area Redevelopment Policies in Britain and the Countries of the Common Market," U.S. Department of Commerce, Washington, D.C., 1965, p. 22.
3. Ohlin, referred to in El Shaks, "Development, Primacy and the Structure of Cities" unpublished M.A. Thesis, Harvard, 1962.
4. Hoover, "Location Theory and the Shoe and Leather Industry," Cambridge, 1937.
5. Isard, W., "Location and Space Economy," M.I.T. Press, 1956.
6. Hoover, op. cit., loc. cit.
7. Skitovsky, T., "Two Concepts of External Economies," J. Pol. Econ. 62: April 1954, pp. 143-151.
8. Davis, Otto, A. and Whinston, Andrew, "Externalities, Welfare and the Theory of Games," J. Pol. Econ., June, 1962, pp. 241-262.
9. Scitovsky, op. cit., p. 145.
10. Meade, J. E., "External Economies and Diseconomies in a Competitive Situation," E.J. March, 1962.
11. Scitovsky, op. cit., loc. cit.
12. Chamberlain, E. H., "Proportionality, Divisibility and Economies of Scale," Q. J. E., 1948.
13. Koopman, T. C., "Three Essays on the State of Economic Science," p. 151.
14. Koopmans, op. cit., p. 152.

15. Berry, B. J. L., "Cities as Systems within Systems of Cities," in Friedmann and Alonso Eds., "Regional Development and Planning," M.I.T. Press, 1964.

16. Koopmans, op. cit., p. 154.

17. Hoover, E. M., "The Location of Economic Activity," McGraw-Hill, New York, 1948.

18. Ohlin, op. cit., p. 219.

19. I am indebted to Schlomo Angel for this illustration.

20. Vernon, Raymond, "Metropolis 1985," p. 110.

21. Vernon, op. cit., p. 102.

22. Mills, Edwin, "Transportation and Patterns of Urban Development," Amer. Econ. Lev., May 1967, pp. 197-210.