

CESIS

Electronic Working Paper Series

Paper No.67

**THE ANALYSIS OF LOCATION, CO-LOCATION AND
URBANISATION ECONOMIES¹**

Börje Johansson² and Ulla Forslund³

May 2006

¹ Status of the paper: The paper has been written as a contribution to a Handbook of Research on Clusters

² CESIS, Centre of Excellence for Science and Innovation Studies, The Royal Institute of Technology, Stockholm and JIBS, Jönköping International Business School. Email: borje.Johansson@ihh.hj.se.

³ JIBS, Jönköping International Business School. Email: ulla.forslund@ihh.hj.se

THE ANALYSIS OF LOCATION, CO-LOCATION AND URBANISATION ECONOMIES

Börje Johansson and Ulla Forslund

Abstract

This paper provides an overview and critical assessment of co-location and clustering in space. Basic location conditions include accessibility to customers, input suppliers including knowledge providers, and regional endowments. A distinction is made between slow and fast location adjustments. In a basic model, it is shown how distance sensitivity and scale economies generate self-reinforcing location externalities. Variations of the same model are employed to illuminate how urbanisation economies can stimulate co-location and clustering. One model variant is designed to examine how innovation activities also can be influenced by urbanisation economies. The paper concludes that a set of basic principals form the basis for localisation and urbanisation economies. However, there remains a challenging gap between model predictions and empirical observations.

JEL classification: L14, L29, O30, R30

Keywords: Location, co-location economies, agglomeration economies, urbanisation economies

1. INTRODUCTION

Location analysis has two major perspectives. The first is concerned with where to place a given economic activity or facility, defined as optimization problem, where the properties of the economic environment are taken as a given fact. The second perspective motivates a different question: how can the entire landscape of activity locations be understood and explained? Both approaches can be associated with an equilibrium framework. However, the second perspective also stimulates the student to think about the evolution of location patterns. How do they emerge and which are the adjustment processes?

1.1 Perspectives on Location Analysis

In the past, economic model-building favoured formulations, where the distance between buyer (receiving agent) and seller (delivering agent) could be ignored. In Debreu's (1959) contribution location indexing is possible, but only in an inessential way. This is manifested in the spatial impossibility theorem attributed to Starret (1978), which implies that the Arrow-Debreu model cannot be applied to a homogenous space, except in the case where transport costs are zero for any spatial competitive equilibrium. As emphasized by Fujita and Thisse (2002), this implies that regional specialization, urban regions and trade cannot be competitive equilibrium solutions.

The above problem has been avoided for a long span of time by introducing different forms of pre-located elements into location models. This is typical for the contribution of Weber (1909), in a model where input suppliers and customers all have a given location and where the single firm selects an optimal location given these facts. This approach remained in the celebrated contributions in the 1950s (Beckmann, 1952; Samuelson, 1952). Likewise, in the contribution of von Thünen (1826), the location of the city and its center is given in advance, and this approach has dominated urban economics since then (Fujita, 1989).

In the subsequent analysis we borrow one fundamental idea from Weber and one from von Thünen. In the latter case a principle idea is that spatial structure is affected by the assumption that the delivery of commodities between seller and buyer is distance sensitive, and this sensitivity varies between commodities. This is as fundamental as it is trivial. The inheritance from Weber is of the same nature, and it amounts to the recognition that a firm as well as a public service provider has to consider both the delivery of inputs and the delivery of outputs. Such an observation is clear cut in Weber's original formulation, but becomes profound when we observe that the location of input suppliers and output customers may change in response to a sequence of individual location choices.

1.2 The Co-Location Phenomenon

The analysis in the subsequent sections is based on the following simple assumption. The economic geography consists primarily of urban regions, some very small and still others very large. Certain transactions are less costly when carried out between actors inside a region than when they take place between actors in different regions. The same applies to certain extra-market information and knowledge flows. In particular, urban regions exist because of these advantages of intraregional interaction.

The subsequent presentation attempts to outline a framework that makes it possible to understand the location decision of a firm in a context of its interaction with its input suppliers and its customers. This should form the basis for modeling co-location

phenomena such as spatial clusters and urban agglomerations. A distinct message is that firms cluster together in a place because they have something in common with regard to (i) resource endowments which are available in the place, (ii) inputs suppliers that are present in the place or (iii) customers that are accessible from the place.

These issues are central in the contribution by Ohlin (1933). He considered four factors as major causes of agglomeration economies. These are

- (i) Internal scale economies
- (ii) Localisation economies
- (iii) Urbanization economies
- (iv) Inter-industry linkages (of input-output type)

According to the framework that is developed in the present paper, only two major causes of agglomeration exist, and these are Ohlin's second and third factor. At the same time, both internal scale economies and interindustry linkages can be essential components of a model that describes localization and urbanisation economies.

Scholars that discuss localisation economies frequently refer to Marshall (1920), sometimes to Hoover (1948) and very rarely to Ohlin (1933). It is as if Ohlin was disqualified because of the Heckscher-Ohlin theorem. The modern reference should definitely be Fujita and Thisse (2002). In addition, their contribution to a deeper understanding of urbanisation economies is equally profound.

In a genealogy perspective Ohlin should be recognised for having brought attention to localisation and urbanisation economies as two parallel agglomeration phenomena, although it is Hoover who has received the credit for doing so. With regard to urbanisation economies the definitions and explanations given by Ohlin and Hoover are quite similar and could be classified as specific to each region, with its given size and composition of a variety of industries. Ohlin and Hoover may differ somewhat in the treatment of localisation economies, where Hoover stresses that localisation economies are industry specific, and this implicitly puts horizontal externalities into the focus. This strong focus on the individual industry is probably unfortunate and is one reason for putting the concept of co-location to the forefront, as is done in the present study.

Co-location may imply clustering of firms that mutually benefit from being located in the proximity of each other, although they do not belong to the same industry. The mutual benefit is a place-specific increasing returns to scale (McCann, 2001). In the analysis developed stepwise in the subsequent sections the major focus is not on horizontal but on vertical externalities. In this context, we attempt in this study to demonstrate that clustering of firms in the same industry primarily is generated by two major forms of clusters:

- (i) *Cluster of input-selling firms.* In this case we observe many firms that supply differentiated and distance-sensitive products to a locally concentrated demand for these inputs. In this general formulation inputs include both service and knowledge deliveries.
- (ii) *Clustering of input-buying firms.* In this case we observe a spatial clustering of firms producing differentiated product varieties in a given place. These firms are attracted to stay in this particular place, because this place has a concentrated supply of distant-sensitive inputs that the firms demand. Thus, the co-location

force is not primarily that the clustered firms produce the same kind of outputs, but that they use similar inputs.

Obviously, these two forms of clusters may combine into a strong pattern of co-location. However, there are clear analytical advantages to be gained by treating them as two individual co-location processes. Moreover, the above approach allows the presentation to continue and depict urbanisation as “a cluster of clusters” and establish an interface with early suggestions in the works of Hall (1959), Artle (1959), Vernon (1962), and later Henderson (1974, 1977). The approach also opens up for a discussion of localisation and urbanisation economies in studies of innovation activities (e.g. Swann, Prevezer and Stout, 1988; Feldman, 1994 and Steiner, 1988).

1.3 Outline of the Presentation

Section 2 presents basic conditions for the location of a firm, such as distance to input suppliers and customers and accessibility to regional endowments. Another important aspect is the distinction between slow and fast adjustments. These ideas are applied in Section 3 that examines how co-location externalities make co-location processes self-reinforcing, where distance sensitivity and scale economies generate the externalities that also comprise R&D processes. Section 3 makes use of a simple model that is used also in Section 4 that deals with urbanisation economies, and where taste for variety is emphasised with regard to both firms’ demand for inputs and households’ consumption demand.. The same model is also applied to illustrate urbanisation economies of innovation activities. Section 5 concludes that it is possible to consider a set of basic principles to describe a firm’s location decisions in the context of localisation and urbanisation economies. This last section also identifies gaps between theoretical suggestions and empirical evidence. In view of this a set of research problems are discussed.

2. LOCATION OF THE FIRM

Section 2 introduces basic elements for an analysis of a firm’s location decisions, such that these elements remain intact in the modelling of co-location phenomena. It is also observed that location decisions have to take into account that certain location characteristics change at a slow pace, whereas others adjust quickly. Finally, it is also stressed that a firm is involved in (i) output activities that can be analysed on a fast time scale and (ii) development activities that have to be depicted on a slower time scale.

2.1 Slow and Fast Location Processes

A classical approach in trade and location analyses is to consider certain resources as pre-located and to examine how other resources or activities adjust their location in response to resources with a given location. For example, network infrastructure for movements of persons and products may be considered as given, whereas the location of economic activities adjust in response to interaction costs, which are influenced by the given infrastructure (e.g. Anderson, 1993). The factor-proportions theory provides another example, where capital and labour are given for each of two regions and cannot move between the regions, and where each region specialises by selecting production activities in concordance with its relative factor abundance.

Many resources have a given location in the short term, but they may shift location in a longer time span. These resources, like built environment and any other located endowments, adjust on a slow time scale. The location of firms and the intensity of

different types of economic activity adjust on a faster time scale, driven by faster adjustment processes. In this sense, short-term location equilibrium is only a temporary phenomenon that applies as long as the slower location processes remain practically unchanged (Johansson, Batten and Casti, 1987).

Figure 2.1 presents three basic conditions that influence a firm’s location decision, including the decision to remain in a given place. The first factor represents the costs of interacting with input suppliers, where interaction costs include the transaction costs that affect the input price (McCann and Shefer, 2004). These costs will vary across alternative locations. In an extended setting interaction also refers to R&D collaboration with input suppliers, including knowledge providers. The location pattern of input suppliers may change more slowly than the activities of the input-buying firm, but the time scales are not necessarily very different.

The second factor represents the costs of interaction with customers. In some cases these are few, and in other cases the firm has many customers that are spread over many regional markets. And such circumstances affect the size of interaction costs, which include transport and other transaction costs. Again, customer-related costs will also depend on where the firm is located. When the firm sells on many different geographical markets, the pattern of customer location will adjust on a slower time scale than what applies for the selling firm itself. Gradual (and path-dependent) changes in the concentration of demand is a core element in the so-called “new economic geography” (Fujita, Krugman and Venables, 1999; Brakman, Garretsen and Marrewijk, 2001).

The third factor in the figure comprises the endowments of the region. Such endowments primarily consists of slowly changing features of a region, e.g. infrastructure and built environment, skills and knowledge intensity of the labour force, as well as amenities in general.

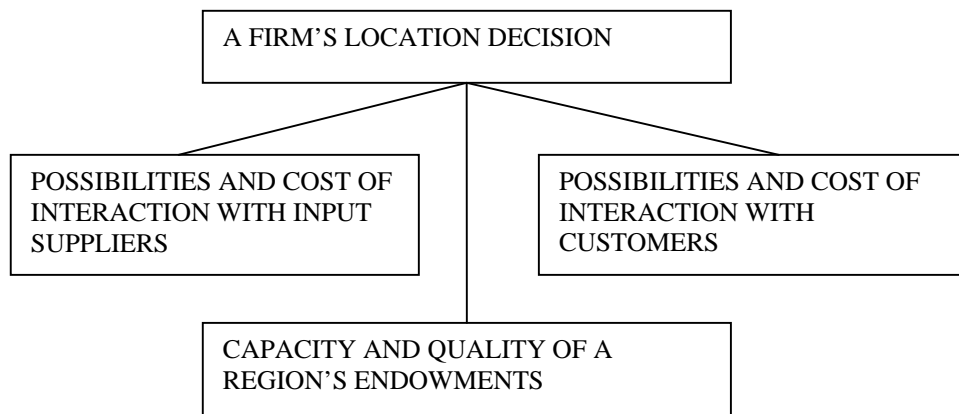


Figure 2.1: Basic conditions that influence a firm’s location decision

In view of the characterisation in Figure 2.1 this section investigates in sequence how a firm’s location can be influenced by its distance to input suppliers and customers, and by the endowments of the region where the firm is located. The presentation avoids to deal with multi-location firms, such as multinationals (Lipsey, 2003). The subsequent presentation in Section 2 primarily deals with general aspects of a firm’s distance to (i) input suppliers, and (ii) customers, where distance is reflected by the firm’s transaction and other interaction costs. In addition, the

presentation considers the benefits the firm can have from the endowments of the functional region in which it is located.

2.2 Distance to Input Suppliers

A firm transforms inputs to outputs that are supplied to customers. The profit of the firm is determined by the firm's revenue minus its input costs. As inputs have to be delivered to the firm, the distance to input suppliers and resource supply points will affect input costs. In a dynamic context the inputs may be related to the firm's development activities, including search for new production routines and new product attributes. In a static setting the decision criterion is to minimise input costs.

The classical Weber problem was presented and solved in Launhardt (1882) and further analysed in Weber (1909). Following the description by Puu (2003), the problem is about the location decision of a firm that uses two different inputs to produce one output. Each input is supplied from a distinct location, which is different for each input. The demand for the output is located in still another location. Thus, everything is pre-located except the place of the producing firm. Typical pre-located input supply-points are mines, oil wells and ports.

The well-known decision criterion for the firm is to select a location such that the sum of transport costs for the two inputs and for the output is minimised. Obviously, the Weber problem can be extended to situations where a firm has several inputs and several delivery points where demand is concentrated (Beckmann and Thisse, 1986).

In Figure 2.2 the Weber problem is illustrated, where transport costs are measured per unit output. The cost of transporting input 1 and 2 to the firm's location in r is given by c_{1r} and c_{2r} , respectively, while c_{rm} represents the cost of bringing the output to the pre-located market (customer). For each location r that the firm selects, the total transport cost per unit output can be calculated as $T_r = c_{1r} + c_{2r} + c_{rm}$. By choosing the location r in an appropriate way the firm can minimise the size of T_r .

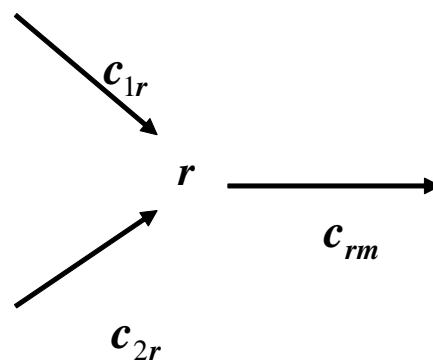


Figure 2.2: Illustration of the Weber problem

The Weber model was later extended by Moses (1958), to consider that firms will substitute inputs for each other in response to relative prices. The result of this extension is that a firm's choice of how inputs should be optimally combined is made in conjunction with the choice of location. Thus, a firm's location problem is embedded in a more general production problem, where the firm minimises the sum of production costs, input costs and output delivery costs (Isard, 1951, 1956).

The Weber location model illustrates for example the choice of location of a steel plant. In this case we may also permit that the production features internal scale economies. However, the model does not have much to say about agglomeration of economic activities in a specific location. In spite of this, the Weber problem remains a core model for the analysis carried out in sections 3 and 4. In these sections we introduce the possibility that the production of input suppliers features internal scale economies, while at the same time some of these inputs are distance sensitive. When the scenery changes in this way the landscape of solutions is altered in a profound way.

The Weber model may also be seen as a basic component in analysing location in a world economy characterised by outsourcing, where the formation of economic networks generates a landscape of subcontracting component production, design and R&D, marketing activities and sales of services that accompany the sales of products (Grossman and Helpman, 2005). In this context it is vital to express transportation and other transaction costs as link-specific costs, c_{rs} , since these link costs are not proportional to geographic distance but reflect transaction arrangements and affinities between each pair of nodes. Such arrangements also include complex so-called supply chains (McCann, 1998, Li and Polenske, 2003).

2.3 Distance to customers

A reversed version of the Weber problem obtains when a firm has several demand points and has to find a production or supply point that can minimise the shipment costs. Extending this formulation to include many supply and demand locations leads to models such as the transportation problem (Koopmans, 1949; Beckmann, 1952), the assignment problem (Koopmans and Beckmann, 1957), and the spatial price-equilibrium (SPE) model (Takayama and Judge, 1971). In all these formulations the demand points are pre-located as well as the locations from which supply is possible. Solving each model generates a flow pattern between supply points and locations where demand is concentrated. Thus, solutions also inform about how much that is delivered from supply points and how much that is purchased in demand points.

An SPE-model can also be extended so that it reflects the location of production capacities, while considering capacity investments in temporal contexts (Takayama, 1994). Moreover, the problem may be formulated such that firms can make “monopoly profits” (Johansson and Westin, 1987), although the standard approach is one of competitive equilibrium. In standard formulations the delivery networks (and the associated costs), the supply functions, and the demand functions are pre-located. Because of this, the solutions are not affected by the spatial impossibility theorem (Fujita and Thisse, 2002:30-60).

The set of models that depict location as a problem of distance or accessibility to customers can be extended to include models of spatial monopoly and oligopoly with varying price-setting rules (Beckmann and Thisse, 1986). Just like the other models mentioned in this subsection, these models are partial, temporary and constrained in various ways to make the problem formulations tractable. As will be demonstrated in section 3, there are two major issues that are critical. The first is about the stability of each particular equilibrium solution and the presence of multiple equilibria. The second issue concerns scale economies in combination with products that are distance sensitive – and for which the sensitivity can vary in size.

2.4 Endowments and the Attractiveness of Regions

Regions have different endowments in the form of climate, natural resources, infrastructure and production capital, skills of the labour force, social capital and many other things. It is a trivial conclusion that existing endowments provide regions with different opportunities or advantages as well as disadvantages. As a consequence the cost of producing a given commodity will vary across regions as a reflection of each region's endowments. The celebrated Heckscher-Ohlin theorem suggests that there is a static equilibrium solution with endogenously determined prices on production factors (endowments) and product prices. However, endowments also influence the development of each region's economy. At any moment in time, endowments are gifts from the past.

What is then a region? In a first step of making the concept precise, the present analysis adopts the view that a region is an urban-like area such that economic agents who are located inside the region have considerably lower costs when making face-to-face contacts with each other as compared with making contacts with agents who are located outside the region. Such a specification comes close to the definition of a functional urban region as elaborated by Cheshire and Gordon (1998). As a consequence, the size of a region matters and has to be considered as an endowment, because a large region has a richer set of potential contacts than a small region. Moreover, for all products that require frequent contacts between buyer and seller, the suppliers of these products have a larger market in a large region than in a small. In this sense the size of intraregional market demand constitutes a regional endowment and it can be assumed to change on a slow time scale.

The second category of endowments comprises various forms of resources that are attached to a region, such as supply of labour with different skills and knowledge, supply of capital, R&D resources and the economic milieu in general (Andersson, Anderstig and Hårsman, 1990; Florida, 2002). Such endowments in a region can be expected to change slowly over time. Moreover, they represent resources or inputs that affect a firm's development activities rather than its normal output activities. Thus, they constitute stimuli to innovations and renewal processes in general.

The endowments discussed above seem to have replaced the traditional endowments in the form of natural resources. But not entirely. With support from Glaeser and Kohlhase (2004), it can be claimed that natural resources in production are far from playing "the first violin" any longer – if ever. However, consumption-related amenities and associated infrastructure constitute an attractor for households to locate in a functional region, and implies that amenities can support the growth a region's local market (Cheshire and Magrini, 2005). This idea may have a long history. Here we can trace it back to Quigley (1990) and Maclellan (1990), who both suggest that the quality of housing and built environment together with attributes of the natural environment affect population growth directly and economic growth indirectly. This dynamic process has been identified as the phenomenon "jobs follow households" (Holmberg, Johansson and Strömquist, 2003).

3. LOCATION AND CO-LOCATION EXTERNALITIES

Section 3 investigates how co-location externalities influence individual firms to locate in a functional region. Such a region may be a small town (or a set of such neighbouring towns) or a large metropolitan region. In the presentation co-location externalities generate localisation economies or, in another terminology, external economies of scale.

3.1 Distance Sensitivity and Scale Economies

The contribution by von Thünen (1826) is recognised as a foundation for understanding the structure of an urban region and the associated land rent curve. We shall emphasise another aspect, by claiming that he detected that the delivery of products from supplier to customer is distance sensitive and that this sensitivity varies across products. For distance-sensitive products, the geographic transaction costs increase sharply as the distance between supplier and customer increases. The importance of von Thünen's finding goes beyond urban economics, because when it is combined with the existence of scale economies proximity externalities arise, and these externalities bring about co-location or clustering forces.

In Johansson and Quigley (2004) it is argued that a seller and a buyer may form a transaction link that can reduce the distance sensitivity of a delivery process. However, link formation is a costly investment process, which means that proximity advantages remain for distance-sensitive products. Moreover, link formation can be assumed less costly when it is carried out by neighbouring firms.

Consider an individual firm, and assume that its production features internal scale economies. If in addition the output of the firm is distance sensitive, the firm will benefit from having a short distance to its customers. This was the basic finding in Krugman (1991), where one group of products can be delivered at zero transaction costs inside the region where the firm is located and with positive costs to another region. The consequence in a two-region model is an output demand externality that can attract labour and production to concentrate in the largest of the regions. The externality depends on a combination of internal scale economies, the size of the distance-sensitive delivery costs and the mobility of production and labour, while observing that the labour generates demand inside the region where it locates. Krugman's simple example is rich in the sense that two regions will both sustain, as long as there is also an industry for which the input supply is trapped in each region, whereas the output from this industry is insensitive to distance.

Let us go back to our individual firm and consider the inputs to its production. Suppose that some or all of these inputs are distance sensitive. Each supplier of such inputs will benefit from a short distance to its customer firm. If in addition the production of these suppliers is characterised by internal scale economies, the results is an input demand externality.

Krugman (1991) discusses but does not model the input demand externality, which arises when one or several firms in a region expand the aggregate output. Such an expansion has the capacity to attract suppliers of distance-sensitive inputs to find a location in a region. When this happens, these suppliers can deliver the inputs at a lower cost and eventually higher quality. Obviously, this type of location response can stimulate the buyers of the inputs to further increase their activity, which implies that a cumulative externality is in operation. The crucial element is the degree of internal scale economies among the input suppliers, combined with distance-sensitive deliveries.

3.2 Co-Location Externalities via the Market

In this subsection we focus on co-location externalities that are generated via the price formation in markets. These price formation processes comprise pure market exchange as well as exchange that is based on transaction links between a buying and a selling firm. The co-location phenomenon takes two forms (that may indeed coexist). The first is co-location or clustering of input suppliers that service the same customer(s). The second is co-location or clustering of firms that demand the same

type of differentiated products as input in their production. In subsection 3.3 the analysis is extended to extra-market externalities.

The initial interest is now to depict a firm that services a demand, D , that is concentrated in a given region. The associated customers may be just one major firm or a set of firms that all purchase the same input(s). The demand may also arise from households in the region. The delimitating factor is that the products delivered (goods and services) are distance sensitive.

The firm's production is depicted by the cost function

$$C(x) = vx + F \quad (3.1)$$

Where vx denotes variable costs and F fixed costs. The demand that faces the firm is modelled by the following function:

$$x = \alpha p^{-\theta} D \quad (3.2)$$

where p denotes price, $\theta > 1$ is the price elasticity of demand, D is the demand indicator (demand budget), and α is a coefficient that reflects the market structure. Suppose that the firm is the sole supplier in the local market. Then the optimal price, \bar{p} , equals (e.g. Brakman, Garretsen and Marrewijk, 2001)

$$\bar{p} = v\sigma, \quad \sigma = \theta/(\theta - 1) \quad (3.3)$$

This price will be viable only if the associated profit, $\pi(\bar{p}) = \bar{p}x - vx - F$, is non-negative, which requires that

$$v\alpha\bar{p}D(\sigma - 1) \geq F \quad (3.4)$$

Formula (3.4) shows that if D is too small relative to F , the firm cannot locate in the region – as long as its sales are limited to the regional market. The customer(s) in the regional market must then buy its distance-sensitive product from extra-regional suppliers at a price, $p^e > \bar{p}$. In such a situation, if D starts to grow, this can trigger the start of a local supplier.

In view of the single firm that we have depicted in (3.1)-(3.3), the condition in (3.4) can be described as an output demand externality that operates via the market. Suppose now that the customer is another firm. For this firm the condition in (3.4) is also an input-demand externality. The simple argument is that if this customer firm can increase its sales, its demand budget, D , will grow, and when that happens it can be served by a local firm A at a reduced price. As a consequence the customer firm (B) can produce at lower costs and eventually further increase its output and thereby increase its demand for inputs. In summary, the model given by (3.1)-(3.4) describes a down-stream externality in view of firm A and an up-stream externality in view of firm B.

Suppose now that the demand budget refers to customers (or just one customer) with taste for variety. This implies that firm A is just one out of a set of suppliers, indexed by j , that are involved in monopolistic competition, as described in Krugman (1980) or Brakman, Garretsen and Marrewijk (2001). If all suppliers of the differentiated

varieties have the same price elasticity θ , each of them will face the following negatively sloping demand function:

$$x_j = p_j^{-\theta} P^{-(1-\theta)} D, \quad P \equiv \left[\sum_k p_k^{1-\theta} \right]^{1/(1-\theta)} \quad (3.5)$$

where P is a price index for the market of differentiated products. Moreover, in this case (3.2) is transformed to (3.5) if we let $\alpha = P^{-(1-\theta)}$ ⁴.

Table 3.1: The clustering of input-buying and input-supplying firms.

CLUSTER OF INPUT BUYERS	When one or several activities in region expand to a sufficient volume, the demand for associated, distance-sensitive inputs will be large enough to attract pertinent input suppliers (with internal scale economies) to the functional region. Such attraction can be generated by one large firm, but may also be generated by several firms in an industry, where the firms supply differentiated products. These firms would constitute an <i>input-buying cluster</i> that benefits from proximity to suppliers of distance-sensitive inputs. Such a cluster is stimulated to grow in response to better quality inputs at prices that fall as demand for them expands.
CLUSTER OF INPUT SUPPLIERS	When one or several activities in region expand to a sufficient volume, the demand for associated, distance-sensitive inputs will be large enough to attract pertinent input suppliers (with internal scale economies) to the functional region. These input suppliers would constitute an <i>input-supplying cluster</i> to the extent that the input suppliers belong to the same industry and supply differentiated inputs. A major mechanism would then be taste for variety among firms in the output industry.

So far the analysis provides some insight to how an input-supplying firm has an incentive to locate in the proximity of an input-buying firm. Two such firms also benefit from a co-location externality as described in Table 3.1. The input-buying firm may be large export firm or several firms with similar input demand. In the latter case they form a cluster of input buyers.

In addition, the supply of inputs may come from many firms that deliver differentiated and distance sensitive input products. In that case they form a cluster of input suppliers.

It should be observed that product differentiation is essential for both the cluster of input-buying and input-supplying firms. The reason is simply that internal economies of scale are assumed to be a general feature of all firms. Thus, without differentiation there would not be many firms. Another observation that we can make in Table 3.1 is that an input-buying cluster is identified with regard to the similarity of input demand among the clustered firms.

3.3 Extra-Market Co-Location Externalities

⁴ The result in (3.5) can in various ways be extended to interregional transactions (Fujita and Thisse, 2002), but at the cost of more constrained assumptions than we want to make here

The analysis here can benefit from the results in the preceding subsection, because intra-market and extra-market externalities are both influenced by proximity. Having said this, it should also be stressed that information and knowledge flows can be based on transaction principles (e.g. purchase of knowledge from a consultant), and collaboration between firms can indeed be supported by explicit and implicit contracts.

However, it is obvious that information and knowledge may be transmitted through flows that take place outside the market, such that the source of the information is not paid by those who benefit from the flow. These flows are labelled spillover flows, and can be conceived as by-products of (i) pure market transactions, (ii) interaction on transaction links, and (iii) interaction in social networks in which economic actors are members.

Table 3.2 presents a simple scheme of spillover externalities. In essence, all these externalities point in a direction away from static analysis towards models of an economy that is constantly changing, although mostly at a slow pace. The reason is simple; in a static world additional information would not have any value. This is directly and indirectly confirmed in contributions such as Echeverri-Carroll (2001), Karlsson and Manduchi (2001), and de Groot, Nijkamp and Acs (2001).

The table reveals that upstream and downstream externalities are more tractable to analyse than are horizontal externalities, where firms producing similar products unintentionally diffuse information to each other. The major problem is that co-located firms in the same industry have a Janus-relation to their competitors. They do not only gain information, they also give away information. If the latter part dominates, that provides a negative incentive for co-location.

Spillover of non-durable information among competing firms may be interpreted as a necessary part of a competitive market (general market information). The consistency problem arises when one confronts the idea of so-called localised knowledge spillovers. This idea has been strongly criticised by Breschi and Lissoni (2001) for mixing knowledge diffusion, locally-constrained public knowledge and tacit knowledge in a completely inconsistent way. This problem needs much more basic research – both theoretical and empirical – and will be examined further in the following subsection.

Table 3.2: Spillover externalities and co-location

Form of externality	Nature of spillover flows and principle mechanisms
VERTICAL	<i>Upstream or input spillover:</i> As a by-product of the interaction

	<p>between an input-buying firm and its input suppliers, information and knowledge can spill over and stimulate incremental adjustments as well as innovations of the buying firm. Co-location or proximity allows the interaction to be communication intensive, and hence the occurrence of distance-sensitive spillover-flows is likely to increase with proximity.</p> <p><i>Downstream or delivery spillover:</i> As a by-product of the interaction between an input-selling firm and its customer firm(s), information and knowledge can spill over and stimulate adjustments and innovations of the selling firm. The effects of co-location are the same as above.</p>
HORIZONTAL	<p>Interaction between firms in the same industry: Due to proximity, competing firms may imitate each other to move towards best-practice solutions and improve these solutions. Competing firms may also collaborate on R&D activities, and this can have spillover effects. The possibility of mutual spillover-benefits among competing firms is greater if the firms supply differentiated products, such that their markets are only partly overlapping.</p>

The spillover phenomena in the table were discussed early by Marshall (1920), and the associated externalities have remained a worry for many economists (e.g. Bohm, 1987). Especially for small functional regions, these phenomena have been identified as clusters or industrial districts, defined as groups of business enterprises and other organisations for which membership within a group is an important element of each individual member firm’s individual competitiveness. Enright (1996, p 191) adds that the conditions binding the cluster together are buyer-supplier relationships, common technologies, common buyers or distribution channels, or common labour pools.

There are two problems with this idea. The first is that a cluster may influence the competitiveness (productivity) of member firms in a static sense, but it may also influence the firms’ capacity to develop and renew their technology over time (Baptista (1996). The second problem is that firms in a cluster can benefit from spillovers due to pure market externalities, network externalities based on contract agreements and social networks in which entrepreneurs and managers are members. So far, theoretical and empirical research has not managed to shed light on how these three factors combine and how they can be separated (Gordon and McCann, 2000).

3.4 Regional R&D Externalities

Firms combine two activities. The first is to use established techniques or routines to produce and supply a given product or product mix. This may be thought of as the firm’s basic “supply activity”. The second area of firm activity can be labelled “development activities”, generally referred to as R&D. In this context a firm makes an innovation by developing new products and/or changing its routines. In its development efforts the firm also relies on information and knowledge flows from its environment and its R&D networks. Some of these flows are transaction based, whereas other flows have a spillover nature.

R&D is a development activity, where the expected output is an innovation. Knowledge is an input to an innovation process, which – if successful – generates new

or improved products and/or new or improved production routines. Knowledge for innovation can schematically be divided into (i) scientific knowledge in the form of basic principles, (ii) technological knowledge in the form of technical solutions, and (iii) entrepreneurial knowledge about product attributes, customer preferences and market conditions, business concepts and the like (Karlsson and Johansson, 2004). Part of such knowledge is placed inside the firm, but a vast share of all knowledge is external to the firm. In its innovation process the individual firm can make efforts to acquire additional knowledge.

Information and knowledge flows as input to a firm's innovation process may arise from market observations, purchase of knowledge from knowledge providers, and collaboration with input suppliers, customer firms, competitors in the same industry, and universities and research institutes. In all these activities additional spillover of knowledge is likely to occur. However, both transaction-based knowledge flows and knowledge spillover are frequently assumed to be distant sensitive (e.g. Acs, 1994; Adams, 2002; Feldman and Audretsch, 1999; Verspagen and Schoenmakers, 2000). A prime argument is that knowledge that can feed R&D and innovation activities is complex and requires a high proportion of face-to-face interaction. As suggested by von Hippel (1994), the information is sticky and hence costly to transmit over longer distances. Another observation is that knowledge diffuses between firms as employees (embodying the knowledge) shift to new employers (Almeida and Kogut, 1999). This latter observation provides additional arguments to focus on local labour markets as approximations of functional urban regions.

A particular area of study is the patent production function, where the patent output often is related to firms' proximity to universities and to other firms that carry out R&D (Varga, 1997, 2002; Hendersson, Jaffe and Trajtenberg, 1995; Adams, 2002; Zucker and Darby, 2005). Studies that examine patent counts have been popular, since patents of individual firms (and of regions) are observable as register data. A recurrent question is: how is patent production affected by proximity to university research and to R&D carried out by other firms (Jaffe, Trajtenberg and Henderson, 1993; Feldman, 1994; Varga, 2001; Fischer, 2001)

Many studies that attempt to verify the importance of proximity to knowledge sources have remained imprecise about the meaning of proximity or have used ad hoc specifications of proximity. Recently, some of these deficiencies have been remedied by making use of Weibull's accessibility measure (Weibull, 1976, 1980). The approach contains two instruments. The first is information about time distances between all zones in an economy. The second is a decomposition of space into three categories: (i) local area, (ii) functional urban region, and (iii) extra-regional area.

First, every R&D activity takes place in some local area, r , and time distances between zones inside such an area makes it possible to calculate the local accessibility, A_r^1 to knowledge sources, K_r , inside the local area, expressed by

$$A_r^1 = \exp\{-\lambda^1 t_{rr}\} K_r \quad (3.6a)$$

where t_{rr} is the average time distance inside the local area, and λ^1 is the time sensitivity for local transport (5-15 minutes) Second, every local area belongs to a functional region that consists of several local areas including the region's major town or city. The accessibility from r to all other local areas, s , in the region, R , to which r belongs is given by

$$A_r^2 = \sum_{s \in R(r)} \exp\{-\lambda^2 t_{rs}\} K_s \quad (3.6b)$$

where $R(r)$ contains all local areas in R except r itself, where t_{rs} is the time distance between r and s (15-50 minutes), λ^2 is the time sensitivity for intraregional transport (contact making), and K_s is the size of the knowledge source in location s . Third, the accessibility from r to knowledge sources in locations, $s \in E(r)$, outside region R is given by

$$A_r^3 = \sum_{s \in E(r)} \exp\{-\lambda^3 t_{rs}\} K_s \quad (3.6c)$$

where λ^3 is the time sensitivity with regard to extra-regional transport, where time distances exceed 50 minutes (Johansson, Klaesson and Olsson, 2003).

The above approach has been applied in Andersson and Karlsson (2004, 2005) and Gråsjö (2005) with data from Sweden, where innovation activities are regressed on observations of A_r^1 , A_r^2 and A_r^3 . In these and similar studies proximity matters. The local accessibility, A_r^1 , has the strongest effect. There is also an effect from intraregional accessibility, A_r^2 , whereas the influence from extra-regional accessibility, A_r^3 , is insignificant. Positive effects can be found for accessibility to university R&D, but if university and industry R&D are included in the same regression equations, the accessibility to industry R&D (carried out by firms) dominates completely. Having said this, there is room for a sincere reservation. It is plausible that R&D-active firms are attracted to regions with university R&D, while at the same time the size of university R&D resources adjust to where R&D intensive firms are located. With this type of interdependent dynamics it matters which of the two processes that adjust fastest.

The above presentation means that a firm can improve its innovation performance by choosing a location that enhances its accessibility to knowledge-intensive labour that can participate in knowledge activities and its accessibility to knowledge sources that can generate knowledge flows that are inputs to innovation activities. In other words, locations with advantageous innovation conditions can attract firms that are innovation intensive and also stimulate firms in such a region to intensify their innovation efforts. These suggestions are summarised in Table 3.3.

Table 3.3: Location of R&D activities

Assumption	Consequence
Knowledge embodied in the labour force of a region and in the R&D organisation of firms and universities is sticky	Then the knowledge assets of a region functions as a regional endowment that attracts R&D active firms to the region.
Knowledge flows between firms, knowledge providers, and other actors influence R&D productivity of firms in a positive way.	As a consequence the location of R&D activities (among firms and universities etc) is a self-reinforcing, cumulative clustering process.
Multi-unit firms like multinationals employ firm-internal channels for knowledge flows. At the same time they benefit from knowledge flows inside the region where each unit is located.	This can generate global R&D and innovation systems that combine network interaction over long distances and intraregional proximity-dependent knowledge flows.
Successful innovation systems are based on a combination of R&D activities and entrepreneurial knowledge.	Innovation activities will concentrate in regions where ordinary interaction between input-supplying and input-buying firms are co-located, where R&D actors are co-located and where entrepreneurship is present.

A crucial element that is not treated in the above discussion is the importance of keeping intellectual capital private. The problem is stressed when the R&D-process in product cycle models is examined. In such a model, firms make innovation efforts because they foresee temporary monopoly profits when new successful (and firm-specific) knowledge is created. Another aspect of the product cycle model is that the innovative firm can benefit from accessibility to diversity of demand, partly as an element of a trial and error interaction between the product developer and pertinent customers (Johansson and Andersson, 1998). Obviously, diverse demand spillovers have a more consistent basis than do the horizontal externalities. Information of customer preferences are transmitted not only for existing but also for products under development. All these observations point towards the phenomenon urbanisation economies, which is the topic in the following section.

4. AGGLOMERATION AND URBANISATION ECONOMIES

In the preceding section the focus is on co-location that brings about localisation economies, which is a form of agglomeration economies. Another form is urbanisation economies, and this phenomenon is based on similar mechanisms but with a focus on the role of diversity.

4.1 Agglomeration and Urbanisation Economies

Urbanisation economies arise in large functional urban regions – in particular in metropolitan regions. In such regions, for which diversity is a basic characteristic, we can observe a variety of specialisations inside the region. In a sense a metropolitan region is a self-contained universe that generates a large share of its own demand. Therefore, in a large urban region there are several location decisions to consider. Should a firm locate in a region in the first place? If the answer is affirmative, in

which local area should it locate? Moreover, does it matter in which zone of a local area the firm is established? The multiplicity of alternatives in a metropolitan region makes these considerations necessary.

Urbanisation and Productivity

When does the selection of a particular zone matter? The general case is co-location of agents that benefit from frequent contacts, and the notion of a central business district (CBD) in the largest city of a region would be the standard case – with the financial district as the prominent CBD example. The other example is the shopping district, where shopkeepers benefit from co-location, because customers benefit from making a single multi-purpose trip to such a district rather than making many diverse trips to complete their shopping. Fujita and Thisse (2002) address this as “the formation of clusters of stores selling similar goods”.

An urban region has several local areas, usually with a largest central city and a set of smaller neighbouring towns. The central city is characterised by the largest diversity of certain contact-intensive activities and by having the highest land values. The local areas may have special accessibility to facilities for long-distance interaction, such as air terminals, harbour services and the like, and they have generically lower land values.

To paraphrase a statement attributed to Ernest Hemmingway, metropolitan regions are different from other regions by being larger and hence capable of according more diversity. In this way they become more attractive to households, to the extent that households give priority to the diversity, which large urban regions can offer. Another assumption, which goes back to Ohlin’s (1933) discussion of urbanisation economies, is that large urban regions are more productive. In several contributions by Henderson (1977, 1988) the suggestion is that this is verified by higher land values in large functional regions. Moreover, a recurrent conclusion in Fujita and Thisse (2002) is that productivity (or efficiency) advantages of agglomerations are ultimately collected as land value incomes in the agglomeration.

The above issue of productivity is discussed in Johansson and Quigley (2004). They separate three mechanisms that enhance productivity:

- (i) An entire industry may benefit from agglomeration, since the size of the agglomeration provides sufficient demand to allow individual firms with internal scale economies to develop differentiated products.
- (ii) An individual firm may benefit from the option to buy more specialised inputs at lower transaction costs from differentiated input suppliers within the region.
- (iii) An individual firm may benefit from knowledge flows outside the market (spillovers) that arise from proximity within an agglomeration.

The first two intra-market cases are symmetrical in the sense that the pertinent externalities are generated simultaneously on the supply and demand side. In the third case the externality is “neutral” in the sense that it may affect all regional agents in the same way.

Urbanisation, Innovations and Growth

Studies of firm and industry location frequently consider the existence of agglomeration externalities as the key force behind clustering. In this context, urban externalities involve diversity of suppliers and information spillovers about market conditions and technology. Models of dynamic externalities argue that clusters grow

because they allow people to interact and learn from each other. The frequency of the interaction is facilitated by proximity. This perspective matters when the analysis is extended from studying productivity to include growth of productivity, such that income per inhabitant is augmented.

Jacobs (1969, 1984) viewed this second type of externality, which emphasises industry diversity in a city or region, as the main factor behind regional development and national economic growth: “People who think of cities simply as towns that have kept growing larger are believers in a *preformation* theory of city growth, an enlargement of what is essentially already there” (Jacobs, 1969: 126). Later work by Bairoch (1988) supports Jacobs’s arguments, particularly with regard to the effects of diversity on innovation and diffusion of technologies. Jacobs’s conclusions were also reiterated by Lucas (1988) in the context of *new growth economics*. Lucas argued that cities play the role of *external human capital* for economic activity and the growth of knowledge. Obviously, these considerations are compatible with how the early phase of a product cycle is described (Vernon, 1966).

Urbanisation economies are those economies of agglomeration, which accrue to firms across different sectors (Jacobs, 1960). People who work in sectors that feature localisation economies will require legal, real estate, retail, educational, health care, and leisure services. Similarly, the firms themselves may require services such as marketing, advertising, catering, packaging, transportation, real estate, and security. These various activities are not directly related to the sector experiencing internal returns to scale and localisation economies. In spite of this they will still cluster in the local economy in order to provide services for the firms and employees of this sector. This clustering is a response to the large market possibilities of an urban region. However, these firms will experience increased local factor prices, which must be compensated by economies of scale if the clustering is to continue. This suggests that urbanisation economies develop because many different sectors are characterised by internal scale economies. As suggested in a previous section, a metropolitan region is a “cluster of clusters”.

However, in the literature localisation economies can appear in a small town as well as in a district of a metropolitan region. Hence, the only distinction made between localisation and urbanisation economies is that the latter phenomenon depends in a fundamental way on the size of an urban region – which can grow large because of urbanisation economies. The subsequent analysis is an attempt to make this conclusion somewhat more precise.

4.2 Taste for Input Diversity

In subsection 3.2 we made use of a simple “prototype model” to illustrate the mechanisms that generate clustering of input-buying firms and of input-selling firms. A related “parable model” can be employed to depict one important aspect of taste for diversity among input-buying firms. To simplify, we will select a case with one input-buying firm and many suppliers of differentiated inputs. In addition, the input-buying firm produces an output under a constant-returns-to-scale regime.

The parable has the following narrative form. Firms and industries that benefit from input diversity experience an upstream externality, given that the inputs are distance sensitive and hence can be purchased at a feasible price only if the inputs are delivered from suppliers in the proximity. The assumption is that firms (or industries) can improve their productivity by using a richer composition of input varieties. To describe this, consider a production function, which relates an output X to the input of labour resources, signified by L , and the input of producer services, Q , where the

value of Q increases as the number, n , of input services increases. Letting S_i denote the amount of services of type i , this yields

$$X = L^a Q^{(1-a)}, \quad Q = \left[\sum_{i=1}^n S_i^\rho \right]^{1/\rho}$$

where $0 < a < 1$, and $0 < \rho < 1$. The Cobb-Douglas production function in (4.1) implies that a given share of the returns is spent on producer services, while X expands as n increases, implying increasing returns. We assume that the service supply is characterised by monopolistic competition. In a regional equilibrium setting this means that a region that has a large output will also have a large diversity of service supply.

Let w be the given price of the L -resource, let p_i be the price of input i , and let p be a given price of the output (possibly exported). We can now see that the total service-input cost is a constant share of sales, such that $\sum_i p_i S_i / pX = (1 - a)$, while $wL / pX = a$. This means that a given share is spent on producer services, while X expands as the number of inputs, n , increases. Thus, a large region with a rich supply of services provides a location advantage.

The externality based on (4.1) can easily be used as an argument in explaining urbanisation economies. As a first alternative, suppose that there are many different industries like the one depicted in (4.1), each with taste for input-variety. Then we can just assume that each of these industries have partially overlapping patterns of demand for inputs. If we in addition consider that a large urban region has many such industries then the number of service varieties, S_i , will be greater in a large than in a small region, and the larger region will be more productive and diversified at the same time. A second alternative is to assume that X represents an output index for many input-buying firms, and then a large region just means a region with a large X .

The possibility to export the output, X , (e.g. a composite export commodity) increases as the size of the input supply, Q , increases. At the same time there is an incentive to increase Q as X increases. In this sense, X remains an exogenous demand factor that stimulates the differentiation of distance-sensitive supply of inputs. This latter part of the circularity is a kind of downstream externality.

The above conclusion may of course be extended, given that the demand for X increases as the export price is decreased. In this case the self-reinforcing effects of the externality can be described by the sequence: (i) increased diversity, (ii) increased productivity and decreased costs, (iii) reduced price and increased export sales, followed by opportunities to further increase diversity. However, a negatively sloping demand for the composite output X would change the model setting and make it more difficult to embed in an equilibrium solution.

To a large extent the discussion in this subsection reflects the, already referred to, ideas of Jacobs about the advantages that are present in a large urban region. However, Jacobs also argues that metropolitan regions are import nodes in the global economy. In this case one may observe that import activities feature both internal and external scale economies, and hence certain large regions like Los Angeles and New York have become dominating import nodes in the USA, just like Stockholm and Helsinki in the small Nordic economies (Andersson and Andersson, 2000). This

would imply that imported goods are both cheaper and more diversified in large urban regions, based on a clustering of importing firms in such regions, and this would increase the diversity of inputs to the production in such large regions. In a subsequent subsection this form of import economies will be considered in relation to urbanisation economies of innovation (Johansson and Westin, 1987).

4.3 Customers' Taste for Diversity

Customers taste for variety generates an output-demand externality on several scales in a large urban region. The first of these scales is related to the formation of shopping districts inside an urban region. The second is related to the formation of the urban region itself and its size.

A particular form of a “specialised shopping district” is a small spatial area in which one can find a cluster of stores selling similar goods, such as spectacles for example. The principles behind the formation of such a place are examined by Fujita and Thisse (2002) with the help of a model in which (i) consumers have a taste for variety and (ii) the sellers (shops) supply differentiated goods or services in a monopolistic competition regime with negatively sloping demand and fixed costs – similar to the assumptions in subsection 3.2. In addition the shopping activity of consumers is distance sensitive. In view of this each firm is aware that consumers will find shopping places with much variety more attractive than those with smaller variety. As the shops cluster into a specific place, this place will be characterised by variety.

The assumption about product differentiation is essential. Differentiation together with taste for variety implies that different shops do not have to fear the competition from the neighbouring firms – as in the competition model of Hotelling (1929). Instead, each firm can sell its variety at a price above marginal cost. As a consequence the firm will be able to compete for land in an attractive spot, i.e., a place where other similar firms also want to be. The negatively sloping demand and associated markup is hence essential for the formation of a shopping district.

First, we should observe that a large urban region is special by having sufficiently many consumers (and other customers) to host many different specialised shopping districts. Second, we also should observe that specialised shopping districts may cluster together into malls and other forms of centres. Indeed, co-located specialised clusters of shops in the central city of an urban region are a basic form of a shopping centre.

The preference function of customers in the above shopping-district model is a variant of the U -function in (4.2), where q_k denotes the amount of product variety that a typical consumer buys, and where n represents the number of varieties. The same function can also be applied to explain the role of diversity for an urban region as a whole.

$$U = \sum_{k=1}^n q_k^\phi, \quad 0 < \phi < 1 \quad (4.2)$$

Consider now that each customer maximises U for a given budget share \hat{m} . With a given number of customers the corresponding aggregate budget is M . The result of a customer's choice can be derived from maximising the Lagrange function

$\Lambda = \sum_k q_k^\phi + \lambda(m - \sum_k p_k q_k)$, where p_k denotes the price of product variety k . The

solution is $p_k = \phi q_k^{\phi-1} / \lambda$. Next, consider that the number of varieties, n , is large enough to make negligible the effect that a change in any p_k may have on λ , which expresses the marginal utility of money (Dixit and Stiglitz, 2004). To the extent that this makes income effects negligible, we can introduce the variable $\theta = 1/(1-\phi) > 1$ to obtain the following demand function:

$$x_k = \alpha_k p_k^{-\theta} M, \quad (4.3)$$

which we recognise from formula (3.2), and where the parameter θ represents the price elasticity and where $\alpha_k = \alpha$ for all products.. The value of α can be thought of as a local approximation, which is constant in a temporary monopolistic-competition equilibrium. As discussed in subsection (3.2), this value of α represents a price index for the market (Fujita, Krugman and Venables (1999).

We shall make the heroic assumption that consumers' demand in an urban region is captured by n demand functions like the one in (4.3). The major merit of the model is that it describes stringently a market in which customers have a strong taste for variety when ϕ has a low value.

Our next assumption is that the urban region hosts n suppliers, each with the cost function that is specified in (3.1), and each supplying one of the varieties included in the preference function. This means that in equilibrium each firm produces just one variety and the output satisfies $x_k = x_0 = (F/v)(\theta-1)$ for every firm. Moreover, the number of products, n^* , and the market price, p , will be

$$\begin{aligned} n^* &= M / F\theta \\ p &= v\sigma \end{aligned} \quad (4.4)$$

Formula (4.4) illustrates the effect that the size of fixed costs, F , has on the achievable diversity. In view of this, observe that when an urban region grows, M will grow, and when that happens, the number of varieties n^* will also grow. As a consequence, the utility of the inhabitants in the region will increase, signalling that the region is becoming more attractive. If consumers can migrate at sufficiently low costs from other regions, a large urban region may continue to grow in self-generating, cumulative process. One basic thing that limits such growth is that competition for space in the region will intensify and land values will increase.

One way to generalise the preference function in (4.2) is to define it as a sub-utility function, U_i , which is a component in an overall preference function

$$U = \prod_{i=1}^K U_i^{\alpha_i}, \quad U_i = \sum_{k=1}^{n_i} q_k^{\phi_i} \quad (4.5)$$

With the specification in (4.5) the consumers' taste for variety is present for many categories commodity groups $i = 1, \dots, K$, each with specific parameter α_i and ϕ_i as well as specific fixed costs F_i .

Consider now that formula (4.5) reflects an essential aspect of urbanisation diversity. Then it is evident that the taste for variety can be better satisfied if also import flows can help to increase the variety. This suggestion is relevant for those

product groups that have sufficiently low distance sensitivity. This may obtain as the result of trade network investments that make the long-haul transportation costs low. Combining this with a well-organised local import and sales organisation will make it possible for importing firms to exploit customer proximity in large urban regions. In this explanation two assumptions are critical. First, large urban regions can be assumed to generate a greater demand for variety than elsewhere. Second, the described import-network solutions are based on costly investments, and these can be motivated only by a sufficiently large demand. The pertinent scale economies can be exploited best in metropolitan regions.

Associated with the outline above is a recent and vast set of contributions, which is sometimes referred to as the “new trade theory”. One reference is the series of early contributions that are collected in Krugman (1990). Other contributions are Helpman (1981, 1984), Brander and Spencer (1984)

4.4 Innovation and Urbanisation Economies

Innovation is the result of a firm’s development efforts. In these efforts knowledge flows from outside the firm play a crucial role. We have earlier stressed information and knowledge exchange with customers, input suppliers and specialised knowledge providers. Metropolitan regions have advantages by having a greater amount and variety of knowledge sources. This was a starting-point in Vernon (1992, 1996) and is elaborated in more recent contributions such as (Feldman and Audretsch, 1999; Acs, 1994, Jaffe A, Trajtenberg M and Henderson R (1993). The so-called Jacobs hypothesis is that metropolitan regions are different by having both intraindustry and interindustry knowledge flows (Capello, 2001). In the subsequent analysis we emphasise that the diversity of industries in a large urban region implies an extraordinary diversity of both customers and input suppliers (including knowledge providers).

To illustrate the features of urbanisation economies in the context of product development, we shall make use of the prototype model introduced in subsection 3.2. This model is altered by introducing a condition that forces the monopolist-competition firms to carry through product innovations repetitively, because established product variants are assumed to become obsolete after a certain time on the market.

Let us consider the following framework. At any point in time, t , a monopolistic-competition market contains a set, $N(t)$, of product varieties. In a growing economy this set can be expected to expand. Each of these products have the same demand function and cost function, as specified in (3.1) and (3.2) – with one important qualification. Each product has a date of “birth”, τ , and the demand coefficient $\alpha = \alpha(t, \tau)$ is a function of its age, $t - \tau$. When the age reaches a certain level, we assume that the α -value drops, which implies that after the drop the production is no longer profitable. This will force firms to develop new products, and when a new product is ready these R&D costs amount to F .

Given the above assumptions, the demand specification takes the form

$$x(t, \tau) = \alpha(t, \tau) p(t, \tau)^{-\theta} D \quad (4.5)$$

where D is a demand indicator (potential demand) for the entire product group, for which all different varieties are included in $N(t)$, where $\alpha(t, \tau) = \alpha$ when the product is introduced and remains at this level until it drops, and where $p(t, \tau) = v\theta/(\theta - 1)$

until demand falls and the supply ends. This happens when $\Delta T = t - \tau$ has passed its critical value ΔT^* . The individual firm has to recover the fixed costs during the time period when it can sell its output at the equilibrium price given above.

In association with the demand given by (4.5) the cost function of the individual firm is given by

$$C(t, \tau) = vx(t, \tau) + F \quad (4.6)$$

and we may consider the advantage the firm has from a location in a large urban region during the life time ΔT^* . We can do that by considering that the input costs of the firm has the following form:

$$v = (\beta\rho + (1 - \beta)\hat{\rho})\bar{v} \quad (4.7)$$

where ρ is the price of locally purchased input, $\hat{\rho}$ is the price of inputs bought from outside the region, where $\beta \leq 1$ is the share of local supply of inputs and $1 - \beta$ is the share of inputs bought externally, and where \bar{v} is a given coefficient. Assume now that $\hat{\rho} > \rho$. Then the firm will benefit from being located in a large urban region, in which the size of β is large. If the production requires input diversity, large urban regions will have a larger set, N , of varieties than a smaller region.

In the setting outlined above, the individual firm has to introduce a new product variety in intervals of length ΔT^* . The associated R&D requires various kinds of knowledge (or R&D) inputs. The cost of acquiring a unit R&D input from inside the region is denoted by ω , whereas a similar input from outside the region costs $\hat{\omega}$. The share of locally supplied R&D inputs is signified by $\delta \leq 1$ and the share of R&D inputs from outside the region is $(1 - \delta)$. Then the fixed cost caused by the product development activity becomes

$$F = [\delta\omega + (1 - \delta)\hat{\omega}]\bar{F} \quad (4.8)$$

where \bar{F} is a given coefficient. Assume now that a large urban region has a greater supply of R&D inputs than other regions. Then we can expect that δ is larger in large than in small urban regions. As a consequence, there will be an R&D advantage for large urban regions, if we assume that $\omega < \hat{\omega}$ due to distance sensitivity of R&D deliveries.

The extremely simple model outlined here manages to suggest two advantages of large urban regions. For a given product group, both the variable and fixed cost components may be smaller. Having said this, we observe that the cost function in (4.6) excludes costs of floorspace, which should be high in large urban regions due to high land values.

However, the prime observation is that in accordance to (4.8) firms in a large urban region have lower fixed costs, as represented by F . Consider now that there are many different products that have the same elasticity θ and the same variable-cost specification as given by (4.7). Then we can say that (4.8) implies that in a large urban region firms can develop new products in product groups for which \bar{F} is high, and this opportunity will not be available in smaller regions.

What conclusions can be suggested from the exercise above? We can formulate three tentative conclusions that should provide incentives to future research. The first

reflects on exports, the second on imports and the third on knowledge flows and intellectual property:

- (i) A firm's development of a new product variety brings about fixed costs. When fixed costs are large the intraregional market is not large enough to allow the firm to recover the fixed costs. As a consequence, a large proportion of all R&D and innovation efforts are linked to export plans. In small regions this is always the case.
- (ii) The analysis above identifies the share β of local variable inputs and the share δ of local knowledge inputs. In a metropolitan region, these shares can be expected to be higher than in smaller regions. At the same time, diversified and large scale import activities also make both the "price" of imported current inputs, \hat{p} , and the "price" of imported knowledge inputs, \hat{w} , lower than in other regions – thus providing an import-based advantage in metropolitan region.
- (iii) The variable \hat{w} is meant to reflect the cost of acquiring a unit of knowledge inputs. If many knowledge flows are not charged any price, this price variable will also be low. This brings us to the point of reflecting on the "knowledge leakage" out from the individual firm, which may be less serious in a large urban region, because the large variety in a metropolitan region could imply that for the single firm the benefits of inflows from a rich environment outweighs the losses outflows that are diffused to firms that – due to diversity – are not specific competitors.

5. IS CO-LOCATION DYNAMICS THE GENERIC PHENOMENON?

In the previous sections a specific set of principles have been employed to depict proximity externalities that constitute one particular reason for differences in regional growth rates and location patterns. However, there is also an empirical issue of finding out to what extent one can attribute regional economic growth to urbanisation economies and what role is played by localisation economies.

The benefits that accrue to a firm due to its location can be traced in several dimensions, such as the sustainability or survival of the firm, its productivity, its innovativeness and its profitability – to mention the most basic economic indicators. Can it be empirically verified that these performance indicators are more favourable for firms located in a spatial cluster or in a large urban region, respectively? There is astonishingly little research that provides conclusive answers to this type of questions.

There are indeed theoretical reasons why empirical analyses may have difficulties to verify cumulative change based on spatial externalities. In some cases an industry cluster may grow, whereas other cases provide examples of stagnation and decline. Obviously, a cumulative change process may come to rest, and possibly satisfy equilibrium conditions. Thus, it is not sufficient with "theory components" that illuminate the cumulative work of externalities. Mature knowledge would include the capacity to depict the growth and decline phases as well as potential equilibria. And empirical analyses should be organised to distinguish and capture change and equilibrium aspects of localisation and urbanisation economies.

5.1 Clustering and Localisation Economies

Consider co-location associated with a specific industry. In this case the following questions may be asked:

- (i) Do firms belonging to a spatial cluster have a higher survival rate than similar firms that are not co-located? When the answer is in the affirmative, it remains to find out the regional conditions that explain the observed sustainability.
- (ii) To what extent is the productivity and profitability higher for firms that are co-located compared to other firms? In addition, does the productivity rise faster for firms that are co-located?
- (iii) Are firms belonging to a spatial cluster more innovative than firms are not co-located?

Presently, the above questions remain to a large extent unanswered. Widening the scope to cover both horizontal and vertical co-location may open for new avenues of empirical research. Existing empirical studies do not allow for conclusions about productivity and growth differentials. However, a given industry cluster may have a past history of cumulative co-location and productivity growth (Audretsch and Feldman, 1995). If this change process has come to rest, one should expect productivity differentials to have disappeared.

In a recent study Beaudry and Breschi (2003) ask the question: are firms in a cluster more innovative? Their answer is clear-cut. In general the firms in a spatial cluster are not more innovative. However, for specific spatial clusters, many firms are innovative and as a consequence they form a milieu from which new entrants into the cluster can benefit. This type of observation suggests that co-located activities survive through repeated product renewal and other successive innovations. Other observations in this concluding subsection tell us that longitudinal perspectives are essential. Moreover, a co-location milieu includes both social networks and policy processes.

5.2 Urbanisation Economies and Dynamics

Urbanisation economies are different from localisation economies precisely because they do not rely on specific vertical and horizontal externalities but on diversity of input alternatives and on customer diversity. This implies that the options of developing urbanisation economies should increase as the size of an urban region grows. The empirical evidence of urbanisation economies is also more clear-cut than in the case of localisation economies. For a range of economic activities one can observe how productivity, wage levels and land rents are higher in large than in small urban regions (e.g. Henderson, 1988).

Recent theory is explicit in its assessment of urbanisation economies. Following Fujita and Thisse (2002), these economies should give rise to productivity gains and higher wages. Moreover, given that the land market is well behaving – and hence is supporting an efficient equilibrium – the surplus from higher productivity will be accumulated in land values, which is reflected by higher floorspace costs. This form of equilibrium would also wipe out profit differentials between similar firms in different regions.

The conclusions about land values are based on models depicting monopolistic competition and taste for variety. This remark also suggests an alternative view on

urbanisation economies, emphasising that the benefit of the inhabitants of an urban region materialises in a greater variety. Students of R&D and innovation also argue that greater diversity in large urban regions generate distance-sensitive knowledge flows, which fosters likewise diverse innovation activities – referred to as the Jacobs hypothesis. This idea is elaborated by Capello (2002), who carries out a detailed study of the Milano region and manages to identify localisation phenomena, for specific innovation-intensive activities, inside the region. Based on a quite different approach Feldman and Francis (2001) arrive at similar conclusions for a high-tech cluster in the U.S. Capitol region.

The authors of the present paper strongly believe that models that focus on externalities that combine horizontal externalities and vertical externalities, with clustering of both input-supplying and input-buying firms can bridge the gap between the localisation and urbanisation routes of analysis. The prime difference being that large urban regions more easily can accord richer combinations of that type, and hence offer more novel options and be more generous in allowing failures.

REFERENCES

- Acs ZJ, ed. (1994), *Regional Innovation, Knowledge and Global Change*, Frances Pinter, London.
- Acs Z, Audretsch DB and Feldman M (1992), Real Effects of Academic Research: Comment, *American Economic Review*, 82:363-367.
- Adams J (2002), Comparative Localization of Academic and Industrial Spillovers, *Journal of Economic Geography*, 2:253-278.
- Almeida P and Kogut B (1999), Localization of Knowledge and the Mobility of Engineers in Regional Networks, *Management Science*, 45:905-917.
- Andersson M and Karlsson C (2004), The Role of Accessibility for the Performance of Regional Innovation Systems, in C Karlsson, S-Å Flensburg and P Hörte (eds), *Knowledge Spillovers and Knowledge Management*, Edward Elgar, Cheltenham.
- Andersson M and Karlsson C (2005), Knowledge Accessibility and Regional Economic Growth, *CESIS Working Paper Series No 40*, Royal Institute of Technology, Stockholm.
- Andersson ÅE (1993), *Economic Structure of the 21st Century*, in ÅE Andersson, D Batten, K Kobayashi and K Yoshikawa (eds), *The Cosmo-Creative Society – Logistical Networks in a Dynamic Economy*, Springer-Verlag, Heidelberg.
- Andersson ÅE, Anderstig C and Hårsman B (1990), Knowledge and Communications Infrastructure, *Regional Science and Urban Economics*, 20:359-376.
- Andersson ÅE and Andersson DE, eds (2000). *Gateways to the Global Economy*, Edward Elgar, Cheltenham.
- Andersson ÅE and Johansson B (2000), The Stockholm Region in the Global Economy, in ÅE Andersson and DE Andersson (eds), *Gateways to the Global Economy*, Edward Elgar, Cheltenham.
- Audretsch DB and Feldman M (1995), Innovative Clusters and the Industrial Life Cycle, *Discussion Paper 1161*, Centre for Economic Policy Research, London.
- Baptista R (1998), Clusters, Innovation and Growth: A Survey of the Literature, in GMP Swann, M Prevezer and D Stout (eds), *The Dynamics of Industrial Clustering: International Comparisons in Computers and Biotechnology*, Oxford University Press, Oxford.

- Beckmann MJ (1952), A Continuous Model of Transportation, *Econometrica*, 20:643-660.
- Beckmann MJ and Thisse J-F (1986), The Location of Production Activities, in P Nijkamp (ed), *Handbook of Regional and Urban Economics – Volume 1, Regional Economics*, North-Holland, Amsterdam, pp 21-95.
- Bohm P (1987), External Economies, *The Palgrave Dictionary of Economics*, The Macmillan Press, London.
- Brakman S, Garretsen H and van Marrewijk C (2001), *An Introduction to Geographical Economics*, Cambridge University Press, Cambridge.
- Brander JA and Spencer BJ (1984), Tariff Protection and Imperfect Competition, in H Kierzkowski (ed), *Monopolistic Competition and International Trade*, Oxford University Press, Oxford.
- Capello R (2002), Entrepreneurship and Spatial Externalities. Theory and Measurement, *Annals of Regional Science*, 36: 387-402.
- Capello R (2001), Urban Innovation and Collective Learning: Theory and Evidence from Five Metropolitan Cities in Europe, in MM Fischer and J Fröhlich (eds), *Knowledge, Complexity and Innovation Systems*, Springer, Berlin.
- Cheshire P and Gordon I (1998), Territorial Competition: Some Lessons for Policy, *Annals of Regional Science*, 32:321-346.
- Cheshire P and Magrini S (2005), Regional Demographic or Economic Dynamism? Different Causes, Different Consequences, Paper presented at a workshop at Jönköping International Business School, June 2005.
- Debreu G (1959), Theory of Value: An Axiomatic Analysis of Economic Equilibrium, Cowles Commission Monograph 17, John Wiley, New York.
- Enright MJ (1996), Regional Clusters and Economic Development: A Research Agenda, in UH Staber, NV Schaefer and B Sharma (eds), *Business Networks: Prospects for Regional Development*, de Gruyter, Berlin.
- Feldman MP (1994), *The Geography of Innovation*, Kluwer, Dordrecht, Boston.
- Feldman MP and Audretsch DB (1999), Innovation in Cities: Science-Based Diversity, Specialisation and Localised Competition, *European Economic Review*, 43:409-429.
- Feldman MP and Francis J (2001), The Entrepreneurial Spark: Individual Agents and the Formation of Innovation Clusters, prepared for the conference on *Complexity and Industrial Clusters – Dynamics and Models in Theory and Practice*, in Milano, June 19 2001, Fondazione Comunita e Innovazione.
- Florida R (2002), *The Rise of the Creative Class and how it's Transforming Work, Leisure, Community and Everyday Life*, Basic Books, New York.
- Forslund UM and Johansson B (2004), Spatial Clusters and Economic Development, mimeo, Jönköping International Business School.
- Fujita M (1989), *Urban Economic Theory – Land Use and City Size*, Cambridge University Press, Cambridge.
- Fujita M, Krugman P and Venables AJ (1999), *The Spatial Economy – Cities , Regions and International Trade*, The MIT Press, Cambridge, Mass.
- Fujita M and Thisse J-F (2002), *Economics of Agglomeration – Cities , Industrial location, and Regional Growth*, Cambridge University Press, Cambridge.
- Glaeser EL and Kohlhase JE (2004), Cities, Regions and the Decline of Transport Costs, *Papers in Regional Science*, 83:197-228.
- Gordon IR and McCann P (2000), Industrial Clusters: Complexes, Agglomeration and/or Social Networks, *Urban Studies*, 37:513-532.

- Grossman GM and Helpman E (2005), Outsourcing in the Global Economy, *The Review of Economic Studies*, 72:135-159.
- Gråsjö U (2005), Accessibility to R&D and Patent Production, *CESIS Working Paper Series No 37*, Royal Institute of Technology, Stockholm.
- Helpman E (1981), International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition: A Chamberlin-Heckscher-Ohlin Approach, *Journal of International Economics*, 11:305-340.
- Helpman E (1984), Increasing Returns, Imperfect Markets and Trade Theory, in R Jones and P Kenan (eds), *Handbook of International Economics*, Vol. 1, North Holland, Amsterdam, 325-365.
- Henderson JV (1974), The Size and Types of Cities, *American Economic Review*, 64:640-656.
- Henderson JV (1977), *Economic Theory and the Cities*, Academic Press, New York.
- Henderson JV (1988), *Urban Development Theory, Fact and Illusion*, Oxford University Press, Oxford.
- Henderson R, Jaffe AB and Trajtenberg M (1995), Universities as a Source of Commercial Technology: A Detailed Analysis of University Patenting 1965-1988, NBER WP 5068, National Bureau of Economic Research.
- Holmberg I, Johansson B and Strömquist U (2003), A Simultaneous Model of Long-Term Regional Job and Population Changes, in ÅE Andersson, B Johansson and WP Anderson (eds), *The Economics of Disappearing Distance*, Ashgate, Aldershot, pp 161-189.
- Hoover EM (1973), *Location Theory and the Shoe and Leather Industries*, Harvard University Press, Cambridge Mass.
- Hoover EM (1948), *The Location of Economic Activity*, McGraw-Hill, New York.
- Hotelling H (1929) Stability in Competition, *Economic Journal*, 39:41-57.
- Isard W (1956), *Location and Space Economy*, Wiley, New York.
- Jaffe A, Trajtenberg M and Henderson R (1993), Geographical Localisation of Knowledge Spillovers as Evidenced by Patent Citations, *Quarterly Journal of Economics*, 108:577-598.
- Johansson B, Batten D and Casti J (1987), Economic Dynamics, Evolution and Structural Adjustments, in D Batten, J Casti and B Johansson (eds), *Economic Evolution and Structural Adjustment*, Lecture Notes in Economics and Mathematical Systems 293, Springer-Verlag, Heidelberg.
- Johansson B and Westin L (1987), Technical Change, Location and Trade, *Papers of the Regional Science Association*, 62:13-25.
- Johansson B and Quigley JM (2004), Agglomeration and Networks in Spatial Economies, *Papers in Regional Science*, 83:165-176.
- Koopmans TC (1949), Optimum utilisation of the Transport System, *Econometrica*, 17:136-146.
- Koopmans TC and Beckmann (1957), Assignment Problems and the Location of Economic Activities, *Econometrica*, 25:53-76.
- Krugman P (1990), *Rethinking International Trade*, MIT Press, Cambridge, Mass.
- Launhardt W (1982), Die Bestimmung des zweckmässigsten Standorts einer gewerblichen Anlage, *Zeitschrift des Vereins deutscher Ingenieure*, 26:106-115.
- Li Y and Polenske KR (2003), Measuring Dispersal Economies, in Entrepreneurship, Spatial Clusters and Inter-Firm Networks, *Research Report 04:01, Uddevalla Symposium*, University of Trollhättan/Uddevalla, pp 615-633.

- Lipsey RE (2003), Foreign Direct Investment and the Operation of Multinational Firms: History and Data, in EK Choi and J Harrigan (eds), *Handbook of International Trade*, 287-319, Blackwell publishing, Oxford.
- Maclennan D (1990), Urban Change Through Environmental Investments, in *Urban Challenges*, SOU 1990:3, Stockholm, pp51-76.
- McCann P (1998), *The Economics of Industrial Location – A Logistics-Costs Approach*, Springer, Berlin.
- McCann P (2001), *Urban and Regional Economics*, Oxford University Press, Oxford.
- McCann P and Shefer D (2004), Location Agglomeration and Infrastructure, *Papers in Regional Science*, 83:177-196.
- Moses LN (1958), Location and the Theory of Production, *Quarterly Journal of Economics*, 78:259-272.
- Puu T (2003), *Mathematical Location and Land Use Theory – An Introduction*, Springer, Berlin.
- Quigley JM (1990), The Quality of Housing, in *Urban Challenges*, SOU 1990:3, Stockholm, pp 39-50.
- Starret D (1978), Market Allocations of Location Choice in a Model with Free Mobility, *Journal of Economic Theory*, 17:21-37.
- Steiner M, ed. (1998), *Clusters and Regional Specialisation. On Geography, Technology and Networks, European Research in Regional Science 8*, Pion, London.
- Swann GMP, Prevezer M and Stout D , eds. (1998) *The Dynamics of Industrial Clustering: International Comparisons in Computers and Biotechnology*, Oxford University Press, Oxford.
- Takayama T and Judge GG (1971), *Spatial and Temporal Allocation Problems*, North-Holland, Amsterdam.
- Takayama T (1994), Thirty Years with Spatial and Intertemporal Economics, *Annals of Regional Science*, 28:305-322.
- Vernon R (1962), *Metropolis 1985*, Harvard University Press, Cambridge Mass.
- Vernon R (1966), International Investment and International Trade in the Product Cycle, *Quarterly Journal of Economics*, 80:190-207.
- Verspagen and Schoenmakers, 2000
- Weber A (1909), *Über den Standorten der Industrie*, JCB Mohr, Tübingen.
- Weibull JW (1976), An Axiomatic Approach to the Measurement of Accessibility, *Regional Science and Urban Economics*, 6:357-379.
- Weibull JW (1980), On the Numerical Measurement of Accessibility, *Environment and Planning A*, 12:53-67
- Zucker LG and Darby MR (2005), Socio-Economic Impact of Nanoscale Science: Initial Results and NanoBank, *NBER Working Paper* 11181.