CHAPTER 4

AGGLOMERATION ECONOMIES

The city is the place where everything affects everything else.
(Werner Hirsch)

4.1 INTRODUCTION

The purpose of this chapter is to analyse the operation and structure of urban areas as well as various factors that may influence the economic potential in an urban area. An urban area should best be positioned to address the issue of economic growth and development due to the concentration of economic activities. Since economic energy is mainly generated in urban areas, an exploration of urban dynamics is important for gaining a perspective on the contribution and role of urban areas to economic growth and development. Public officials, urban and regional planners, economists and the public in general are concerned about ways in which economic growth in their cities can be enhanced.

Due to the characteristics of a densely populated urban area, certain potential economic opportunities exist in these areas. Various concepts will be analysed to emphasise the growth opportunities in an urban area. Internal economies of scale in production allow firms to produce goods more efficiently than individual members. The principle of comparative advantage fosters trade and the development of cities. Agglomeration economies in production cause firms to cluster in cities and this clustering also causes economic growth and development in cities. The principles of both internal economies of scale and comparative advantage are part of the concept of agglomeration economies. A brief reference to the first two concepts will be made, with the remainder of the chapter being devoted to the concept of agglomeration economies in explaining economic growth in urban areas.

4.2   AGGLOMERATION ECONOMIES

Agglomeration economies is one of the central concepts in urban economics. According to this notion, cost reductions occur because economic activities are located in one place.   The existence of development nodes or clusters is explained by agglomeration economies. In part, agglomeration economies mean the advantages of spatial concentration resulting from scale economies. It also refers to the advantages of spatial concentration due to the scale of an entire urban area, but not from the scale of a particular firm (Mills & Hamilton, 1994: 20). By locating close to one another, firms can produce at lower costs. This is an example of a positive externality of production. The production cost of a particular firm decreases as the production of other firms increases. The existence of agglomeration economies is fundamental to the economic explanation of urban growth.

Economies of scale are very much part of agglomeration economies of scale and will be explained briefly. The extent of economies of scale (the amount by which unit costs fall as production is increased) may vary greatly between various production activities (Mills & Hamilton, 1994: 10). Economies of scale are crucial to the existence of urban areas. In the absence of economies of scale, goods and services could be produced on an arbitrarily small scale to satisfy the demand of small groups of consumers (Mills & Hamilton, 1994: 9).   The combination of economies of scale and transportation motivates producers and consumers to locate near production facilities that are large enough to satisfy the demand in the surrounding area. Goods and services can be produced more efficiently on a large scale than on an arbitrarily small scale.

Scale economies arise due to two reasons. Firstly, factor specialisation increases productivity because a worker's skills increase with repetition and workers spend less time switching from one task to another. Secondly, indivisible inputs have a minimum scale of efficiency. If an indivisible input is cut in half, the total output of the two halves is less than the output of the whole. As output increases, the firm uses increased amounts of indivisible inputs, thereby increasing productivity (O'Sullivan, 1996: 20).

If workers receive a wage that makes them indifferent to whether they work at home or at a factory, a small urban area develops around a factory. The workers live near the factory to save on commuting costs and thereby increase the value of land near the factory. To economise on land the workers occupy smaller lots of land and the population density around the factory, relative to the rest of the region, increases. Since a city is defined as an area with a high population density, the factory brings about the development of a small factory-city. The factory-city develops because workers can specialise and use their wages to buy other necessities, and economies of scale are large enough to underpin prices that workers would have asked had they been working and selling from their homes (O'Sullivan, 1996: 21).

As will be seen, the principle of economies of scale is also linked to the concept of comparative advantage. An urban area needs to identify and strengthen its comparative advantage in order to distinguish itself from other urban areas. There are three dimensions that will often determine an area's comparative advantage, its demand, its production, etc. These dimensions are depth (quality of an area's environment and ingenuity of its people), diversity of its economy, and scale of activities (Hirsch, 1973: 177).

The concept of comparative advantage can be explained briefly as follows. Assume that the productivity between two separate regions differs. The difference in productivity could be caused, inter alia, by differences in labour skills, weather or soil quantity, etc. This difference in productivity will ensure that some regions can produce certain goods and services cheaper than other regions. Because prices differ between regions, some regions may start trading various products. The location decisions of traders cause the development of market cities. People employed by trading firms tend to locate near the marketplace to save on commuting costs resulting in higher land prices. People then opt for smaller pieces of land and thus increase the population density relative to the surrounding area. A city now develops due to the high population density. The difference between the productivity that generates comparative advantage is large enough to offset transportation costs, so trade occurs (O'Sullivan, 1996: 18).

It is important to turn to agglomeration economies and explain and anlyse this concept in further detail. Assume that a production function for the manufacturing sector of an urban area can be specified as follows:

$$Q = A(z,t)F(K,L) \qquad\qquad (4.1)$$

In this equation z is a collection of factors that create agglomeration economies, t the level of technology and F(K,L) the function of capital and labour (McDonald, 1997: 339). There is, however, a critical distinction between agglomeration economies and technological progress. Technological progress implies that a firm applying new technology can produce more output with a given amount of capital and labour. Technological progress is therefore seen as internal to the individual firm. Agglomeration economies, by contrast, are external to the individual firm in an urban area. However, it may be that agglomeration economies act by increasing the rate of technological change for firms in a given urban area. Thus, there may be agglomeration economies in the invention and development of new technologies and there may be agglomeration economies in the adoption of new technologies.

It is now important to classify different types of agglomeration economies. A distinction within agglomeration economies is that they may be static or dynamic. Static agglomeration economies mean that the level of some agglomerative factor may be associated with some level of output from industry. In the production function (4.1) above, the level of z creates a level for A(z,t) and hence for output. In this example, a larger urban area may have a better and cheaper form of air transport (the z factor). This will create a once-off increase in A(z,t) and thus a once-off decrease (a level effect) in the industry's cost curves. By contrast, a dynamic agglomeration economy means that the level of the agglomerative factor is associated with a continuous increase in the output of the industry. More inventive innovators will be available in a large urban area, which will create a continuous flow of technological change (growth effect) that exceeds that of smaller areas. The size of the urban area (the z factor) causes technology (the t

factor) to increase continuously with a positive effect on economic growth (McDonald, 1997: 340).

The other major distinction within agglomeration economies is between localisation and urbanisation economies. Localisation economies of scale, occur when firms benefit from being close to other related firms. Urbanisation economies of scale, occur when firms benefit from being located in a large city, even though its firms may be unrelated (Bogart, 1998: 12). Both types can be static or dynamic. The clustering of activities will unfortunately lead to congestion and related negative externalities. The marginal benefits of clustering should at least exceed the marginal cost to some degree, or else the formation of cities will never be observed. The two types of agglomeration economies will now be discussed in more detail.

4.2.1 Localisation economies

Localisation economies occur when production costs of an entire group of firms in a particular industry decrease as the total output of that industry increases at that location (O'Sullivan, 1996: 24). Localisation economies generate clusters of firms in the same industry. Firms in the cluster exploit scale economies in the production of specialised inputs, by sharing the suppliers of these inputs. The cluster attracts not only the demanders of intermediate inputs, but also the suppliers. Firms also exploit scale economies in the provision of specialised business services and local public services. Urban planners may develop a strategy designed to create a cluster of closely related firms in order to attract further growth. The sources of localisation economies will now be discussed.

4.2.1.1   Pool of labour and knowledge

Related firms, locating close together, may contribute to the development of a skilled labour pool. If firms face unstable labour demands, the labour market advantages culminating from agglomeration are particularly useful. Firms can then expand their

workforce quickly due to the large amount of qualified workers available. Fluctuations in employment may increase the advantage of a concentration of skilled labour. A shortage of skilled workers for a particular industry may be addressed by developing a school or training college to improve the quality and availability of labour.

The notion of localisation may contribute to employment creation because of an increased demand for and thus concentration of employment. Employment growth in the local industry can be divided into three parts: The first is that which can be attributed to total employment growth in the country. Next, a certain percentage may be attributed to the fact that national employment growth in a particular industry was more rapid (or slower) than in the country as a whole. The third part is where a comparison is drawn between the employment growth in the local industry and that industry's national growth rate. Here, a direct comparison is made between the industry's growth locally and nationally and is therefore normally called the competitive position of the local industry (McDonald, 1997: 359). A possible test for dynamic agglomeration economies is whether an industry is growing more rapidly in the local economy than in the national economy.

At the centre of dynamic agglomeration economies is the production and use of knowledge. External economies arising from knowledge spillovers are critical to the level of productivity or the rate of economic growth in a country. The question is whether knowledge spillovers come primarily from firms within one's own industry or from firms in other industries. From a static point of view, localisation economies can stem from labour specialisation, better training or learning about the most efficient production process from other firms in an urban area. In a dynamic sense, however, a mechanism for continued reduction in costs is necessary. The greater concentration of firms in a particular industry in an urban area will cause a greater rate of new product development, improvement in existing products, as well as improvements in the methods of these products (McDonald, 1997: 344). The rate of innovation will increase if more highly trained people engage in trying to improve the industry. The information may be transmitted through highly trained workers who move from firm to firm, or

71

through business meetings and conferences, industrial spying, copying of competitor's products and improved training programmes.

In modern times, due to the availability of the Internet, the question may be asked whether innovators and imitators have to be located in close proximity. In certain cases they do not, but in some of the most important cases close proximity is necessary. Mills (1992) has called this situation the transmitting of ambiguous information. He defines ambiguous information as "information that requires an interactive and convergent set of exchanges before the final exchange can be consummated" (Mills, 1992: 11). In the case of an industrial buyer and seller of a specialised piece of electronic equipment, a series of meetings between specialists in design, production, marketing and other departments is necessary before the contract can be signed. In the same way innovators who live in close proximity are more productive than if they are isolated because they communicate interactively. Knowledge spillovers also take place across different industries. A diversity of industries may be more stimulating to the production of new ideas than the size of an individual industry. Sometimes, it is diversity rather than uniformity that delivers new products and new technologies (McDonald, 1997: 345). The knowledge spillover is thus wider than in an individual industry but not as wide as the entire urban economy.

4.2.1.2   Market structure

The central question here is whether a monopoly (or oligopoly) nurtures technological change because it can afford research and development, or is such change due to competitive industries seeking a competitive edge?

Schumpeter (1942: 32) states the following:

> *Possibilities of gains to be reaped by producing new things or by producing old things more cheaply are constantly materialising and calling for new investments. These new products and new methods*

*compete with the old products and old methods not on equal terms but
at a decisive advantage that may mean death to the latter.*

He believes that the existence of large firms would increase the rate of product and process innovation. Galbraith (1956) argued that an oligopoly is the natural outcome of industries in which firms have reasonable economies of scale. He continues that these firms charge excessive prices and engage in wasteful advertising and product differentiation but also produce socially beneficial technological progress. Galbraith (1956: 88) states that:

*the net of all this is that there must be some element of monopoly in an
industry if it is to be progressive.*

The adoption of new technologies tends to be associated with larger firms, lower cost innovations, more flexible management and more complete information.

By contrast there is also a belief that competition will foster innovation. The feeling is that stiff competition leads to the creation and adoption of innovations. A competitive market structure has more entrepreneurs; people who are willing and able to take the risk of starting up new businesses. An urban area with competitive industries is likely to create new businesses and more growth. Financial institutions should thus be prepared to deal with smaller borrowers and should be more receptive to the entrepreneur.

It is generally agreed that both the development and adoption of new technology depends on (McDonald, 1997:347):

i)   Appropriability (ability to capture the benefits);
ii)   market structure; and
iii)   technological opportunity.

It seems that there is no apparent association with the size of the urban area.

4.2.1.3   Specialised machinery

Another source of localisation economies is the ability to share specialised machinery and other production factors. If, for example, enough firms locate together, developing a distribution and warehouse centre, a large enough volume of activity may develop. This may lead to the establishment of a distribution equipment firm that sells, produces or modifies loading and handling equipment. This specialised distribution equipment firm could not have been established in an area where only one or two firms would have need of such equipment. As a result, all firms in the area operate more efficiently.

4.2.1.4   Imitation, modification and innovation

Firms locating close together may be able to copy and imitate one another more readily. Therefore they may respond more quickly to changes in their industry than if they were located farther from their competitors. Although the firm that is copied may be harmed in the process, the cluster of firms locating together may experience a benefit as a whole. A firm that copies a certain change from another firm may be in a better position to innovate even further. These knowledge spillover effects tend to be an important source of localisation economies (Blair, 1995: 99).

4.2.1.5   Shopping externalities

A shopping externality occurs if the sales of one store are affected by the location of other stores. If both stores experience an increase in sales, shopping externalities are generated. That means that each store attracts consumers to the cluster, generating benefits for the other store as well. There are two types of products that generate shopping externalities, *viz.* imperfect substitutes and complements. In the case of imperfect substitutes, the clustering of firms decreases shopping costs and attracts potential buyers. Clustering also occurs when firms sell complementary goods. These type of goods are often purchased on the same shopping trip. Retail clusters provide a mix of imperfect substitutes (shoe stores) and complements (food and liquor stores),

allowing both comparison shopping and one-stop shopping (O'Sullivan, 1996: 32). Retailers that choose to locate in isolated areas instead of clusters, sell goods that are not necessarily subject to shopping externalities.

The desire of individuals to compare similar products like shoes, is an example of comparison-shopping. The agglomeration of a few shoe stores in the same mall may benefit all the shoe stores because it is a more desirable place to shop for shoes. Although one new shoe store in the mall may lower the percentage of shoes purchased at each existing shop, the total sales may increase due to a greater number of consumers. The shopping costs of households may be reduced but retailers, due to more sales, may capture some of these advantages. The owner/s of the shopping centre may also benefit because they can charge retailers higher rents due to the popularity of the shopping centre.

This display-variety agglomeration will occur in cases where products are differentiated with price variations for comparison-shopping. Automobile and shoe stores are examples of agglomerations based on display variety. Complementary products also tend to cluster, although they do not necessarily fall in the same industry. An example of this may be a theatre and a restaurant locating together. Agglomeration clusters may have similar outputs, similar production techniques (but different outputs), or similar input requirements (Blair, 1995: 100).

4.2.1.6   Internal agglomeration economies

Internal agglomeration economies are cost reductions per unit that accrue to a firm that expands its plant in a particular area. The firm receives the benefit of this expansion and therefore the agglomeration economies are internal. In this case the fixed costs are spread over a larger output. Other sources of internal agglomeration economies include division of labour, use of alternative technologies and savings through bulk purchases.

4.2.1.7   Linkages

One of the most important causes of industrial agglomeration is firms trading with one another and therefore locating in the same area. Inter-industry agglomeration occurs through backward and forward linkages. A forward linkage is where suppliers would attract buyers and a backward linkage is where buyers attract suppliers. Less developed countries (and less developed regions) are characterised by weak interdependencies and linkages. Low levels of trade occur between firms in this instance. Primary goods - normally produced by less developed countries – are often exported without encouraging additional local economic activity. Ineffective or absent linkages leads to the inability to generate further growth (Blair, 1995: 96).

4.2.2 Urbanisation economies

Urbanisation economies are cost savings that accrue to firms when the volume of economic activity in an entire urban area increases. The firms in this case may be unrelated. Urbanisation economies differ from localisation economies in two ways: Firstly, urbanisation economies result from the scale of the entire urban economy and not just the scale of a particular industry. Secondly, urbanisation economies generate benefits for firms throughout the city and not just firms in a particular industry (O'Sullivan, 1996: 28). The different sources of urbanisation economies will now be analysed.

4.2.2.1   Infrastructure

Urban infrastructure can be classified as roads, sewers, fire protection as well as recreation and health facilities. Urbanisation economies may result from economies of scale in public infrastructure. Infrastructure is an important input in a diversity of private production and consumption. If the standard and quality of infrastructure provision is on a high level, an increase in the size of an urban area allows lower per unit infrastructure

costs (Blair, 1995: 101). These savings in costs may be passed on to producers or consumers in the form of lower taxes.

The transport sector is one of the main components of an urban infrastructure. The larger the amount of firms, the higher the quality of transport facilities is likely to be. Firms using transport facilities will locate near these transportation nodes with the resulting benefits to these firms.

4.2.2.2   Division of labour

Urbanisation economies may result from a more extensive division of labour due to the greater size and activity of the urban area. In a relatively small urban area, many aspects of production and distribution must be carried out within the same plant because of a lack of specialised firms. Certain activities would therefore be purchased elsewhere or not be carried out at all. The extra costs of obtaining these goods will reduce the firm's competitive advantage relative to other firms.

4.2.2.3   Internal economies of scale

Firms selling to various other firms and households may achieve cost reductions as an urban area expands due to internal economies of scale. Internal economies of scale may be passed on to consumers or production factors.

4.2.2.4   Averaging of random variations

Larger urban markets allow for an averaging of variations in economic activity. If a decrease in the quantity demanded by one customer or group of customers is experienced, it can be offset by an increase in the demand from other customers.

Mills and Hamilton (1984: 18) summarise this aspect of agglomeration economies as follows:

> *The most important of such agglomeration economies is statistical in nature and is an application of the law of large numbers. Sales of outputs and purchases of inputs fluctuate ......for random, seasonal, cyclical and secular reasons.*

A firm within a large urban area would therefore experience less scheduling production problems than a firm located within a much smaller urban area. Labour changes can also be accommodated with more ease within a large urban area than within a small town.

4.2.2.5   Urban diseconomies

As the economic concentration increases, the presence of diseconomies starts to becomes more relevant. Urbanisation economies may, to a certain extent, be partially offset by urban diseconomies (Blair, 1995: 102). Some social scientists are of the opinion that crime, anxiety and loneliness are personal costs involved in high density areas. Urban diseconomies may take the form of inconvenience and delay associated with congestion. Due to the high level of competition among firms to locate near large agglomerations, rents increase. Higher wages may also be seen as compensation paid for the inconvenience of working in congested areas. However, the productivity generally increases as the size of the urban area increases. Therefore, urbanisation economies tend to outweigh urban diseconomies over the range of the city size (Blair, 1995: 102).

4.2.2.6   Empirical estimates of agglomeration economies

One approach to measuring agglomerative economies is to estimate the effects of changes in industry output and city size on labour productivity.

The following equation may be used (O'Sullivan, 1996: 28):

$$q = f(k, e, Q, N) \tag{4.2}$$

where

q = output per worker in a particular industry

k = capital equipment per worker

e = education level of workers (a measure of labour skills and productivity)

Q = total output of the industry

N = total population of the urban area

Output per worker (q) should increase with capital per worker (k) as well as the education level of workers (e). In the case of localisation economies, output per worker also increase with Q (industry output). In the case of urban economies, output per worker increases with N (population). With this relationship it is possible to estimate the independent effect of changes in Q (industry output) on output per worker. This is basically the increase in output per worker per unit change in Q, holding all other factors of labour productivity (k, e, and N) constant. It is also possible to estimate the independent effect of changes in N (city size) on output per worker. From equation (4.2) it is possible to see that education levels (human capital) form an important ingredient in measuring agglomerative economies. One should expect a higher number of qualified people in urban areas than in the surrounding rural areas.

It is now important to enumerate the basic economic functions performed by urban areas. The next section provides a comprehensive and realistic method for classifying urban areas according to their economic functions.

4.3 INDUSTRY CLUSTERS

Noyelle and Stanback (1983) have proposed a scheme for the enumeration and measurement of the fundamental economic functions performed in an urban area. They

grouped industries (as defined by the Standard Industrial Classification (SIC) code system) into eight basic functional areas.

Table 4.1   Basic economic functions performed by urban areas.

1. Agriculture, extractive, construction
• Agriculture
• Mining
• Construction
2. Manufacturing
3. Distributive services
• Transportation, communication and utilities
• Wholesale trade
4. Complex of corporate activities
• Central administrative offices
• Finance, insurance and real estate
• Corporate services
• Business services
• Legal services
• General professional services
• Social services
5. Non-profit services
• Health
• Education
6. Retailing
7. Consumer services
• Hotels, etc.
• Auto repair, garages
• General repair services
• Motion pictures
• Recreation services
• Private household services
8. Government and government enterprises

Source: Noyelle. & Stanback. 1983.

The first category (agriculture, extractive and construction) includes industries not necessarily linked to an urban character, although certain urban areas may specialise in one of these industries. The other seven functional areas are typically urban in character. Manufacturing of goods and the distribution of goods and services forms the next group. The following group is called the complex of corporate activities. This important urban sector includes administrative offices like finance, insurance as well as corporate services such as legal, social and other professional business services. The remaining basic economic functions are performed by the other sectors such as health and education (non-profit), retailing, consumer services and government sectors. In Table 4.1 a summary of the method of classification by type of output is shown.

To determine whether industries do cluster together in urban areas, a location quotient may be used. A location quotient is a measure of the extent to which an urban area specialises in a particular industry. It is defined as the percentage of total employment in an urban area engaged in a particular industry, divided by the corresponding percentage for the nation as a whole.

In algebraic form, the location quotient can be expressed as follows:

$$LQ = (e_i/e)/(E_i/E) \qquad (4.3)$$

where e_i is the employment in industry i in the subject urban area, e the total employment in the subject urban area, E_i the employment in industry i in a specific country and E the total employment in that same country. A location quotient that is greater than 1 indicates that the urban area is probably producing the good or service for export outside its own area. A location quotient smaller than 1 suggests the area is probably importing the good or service (McDonald, 1997: 65).

Research done by Henderson (1986) showed that industries subject to localisation economies tend to cluster in a relatively small number of urban areas rather than being spread out over a large number of urban areas. The general absence of urbanisation

economies in manufacturing industries means that these firms have no particular incentive to locate in the largest areas. Economies that arise through inter-industry linkages mean that industries with strong input-output linkages will cluster together (McDonald, 1997: 65).

4.4 SUMMARY

Urban areas are generally defined as areas with high levels of population density. This concentration of people causes certain economic side effects that may be embodied in the concept of agglomeration economies. Agglomeration economies lead to general cost reductions due to spatial concentration of economic activity. Firms trading with one another normally benefit from locating close to each other. The sources of localisation economies can broadly be divided into three major aspects. The first source is the benefit of labour pooling, including access to specialised labour skills for firms and access to a variety of employment opportunities for workers. The second source is the benefit that stems from economies of scale in intermediate inputs for a product. Lastly, the greater ease of communication made possible by proximity to competitors, suppliers and customers is also a benefit. This includes the ability to pass innovations along rapidly.

There are also three major sources of urbanisation economies of scale. The first source is access to a larger market that reduces the need to transport products over long distances. Secondly, the easy access to a wide variety of specialised services is more readily available in larger cities than small ones. Lastly, there is greater potential for cross-industry spillovers of knowledge and technology.

Location quotients are a popular method for comparing the size of a local industry to that industry's importance in the national economy. The advantage of using this quotient is its inexpensive nature and the fact that it can be applied to both goods and services.

This chapter has shown that the agglomeration economies generated in urban areas are conducive to opportunities for economic growth and development. The mere fact that firms are located within close range ensures that they are well positioned to exploit the benefits provided by proximity within urban areas. The sources of localisation and urbanisation economies are contributing to and strengthening the economic possibilities of this spatial concentration. However, the concentration of economic activities and people that are responsible for creating positive externalities are also responsible for the generation of negative externalities that could offset the expected economic outcome. It is therefore important to address these issues to limit the negative effect thereof on the urban economy. The first of these issues to be addressed is the notion of land use within urban areas.

A wide variety of land use opportunities exist within an urban area. Land use in an urban area should be optimally utilised because of the limited amount of land available. These different land uses will be discussed in the next chapter.