2 Agglomeration economics

Philip McCann

1 Introduction

Agglomeration economics and the economics of industrial clustering have emerged over the last two decades as central issues of research into economic growth and performance. For both urban and regional economists and also economic geographers, the increased interest in these topics is obviously very welcome indeed. However, as the quantity and variety of research in this field has burgeoned in economics, economic geography, urban planning and even sociology, the mixing, and sometimes rather liberal use, of technical terminology, has led to a certain amount of analytical overlapping, confusion and duplication. For this reason, this chapter will consider the origins and analytical foundations of different views and hypotheses regarding the potential advantages of agglomeration, the structural assumptions underlying agglomeration and clustering, and finally the empirical challenges associated with these issues. The aim of the chapter is therefore to disentangle these various issues and to clarify the analytical foundations of agglomeration economics, as well as to outline the empirical challenges associated with these issues. In order to do this we adopt a transactions-costs framework which has been employed elsewhere. The reason is that this is the most parsimonious way of analytically disentangling these various issues from a range of different approaches and different analytical traditions. As we will see in this chapter, while analytically classifying different types of agglomeration and clusters is itself difficult, the empirical identification of these is also very difficult. As such, the observation and measurement of agglomerations and clusters is a topic which, although it initially appears to be rather straightforward, is actually a topic which requires very careful theoretical and empirical analysis.

2 The importance of clustering and agglomeration

A generally observed phenomenon is that most industrial and commercial activities tend to be clustered together in space. These clusters may take the form of industrial parks, small towns or major cities, and most countries generally exhibit a size and activity distribution of cities (Gabaix, 1999a,b), with different ranges of activities taking place in different centres. This observation, however, raises the obvious question of why it is that activities are generally grouped together geographically and what role this clustering plays in economic growth. How does such industrial clustering affect the economy; are there any dynamic aspects to this clustering; and how rapid are any such changes in the urban economic system? These are all key questions for urban and regional economists and economic geographers, yet the importance of these questions has only really become obvious to other economists over recent years.

Since the end of the 1980s and the beginning of the 1990s there has been a major change in emphasis in regional growth analysis. Prior to this, traditional neoclassical regional growth models (Borts & Stein, 1964) were developed on the basis of the original macro-growth models of Solow (1956) and Swan (1956). These early regional models, and also

their more recent subsequent manifestations (Barro & Sala-i-Martin, 1992), were essentially aspatial in construction. Geography was not explicitly modelled in this framework, with the analytical focus being primarily on the interregional factor allocation conditions. Such models incorporated geography only implicitly in their logic by assuming that, as long as markets were competitive and factors were allocated efficiently, then the factor migration process implicit in the allocation process would itself maximize both the levels of technology and also its interregional diffusion.

As such, technological diffusion across space was assumed to be both maximized and equalized simply by the existence and operation of competitive markets. Economic convergence (Barro & Sala-i-Martin, 1992) across geographical spaces was assumed to be the natural outcome of these one-sector model approaches. At the same time, even heterodox models of regional growth (Dixon & Thirlwall, 1975; McCombie & Thirlwall, 1994) which pointed rather more to divergence than convergence, also had no explicit geographical dimension to them. On the other hand, since the early 1990s, the more recent analytical and empirical approaches to regional growth have tended to be explicitly spatial in their construction, focusing on the types of characteristics of different locations which may contribute to localized growth. These more recent approaches do not assume that interregional technology flows are necessarily maximized by simple factor allocation processes; rather, they consider the conditions under which such flows may or may not occur within a competitive environment. In particular, these more modern regional growth analyses tend to emphasize the role played by agglomeration economies and industrial clustering in promoting both specifically local, as well as national, economic growth. This shift in analytical focus has come about for several reasons: one reason is analytical, the second reason is technological, and a third reason is institutional.

The first reason why cities and industrial clustering have become a central focus of research in economic growth analysis is primarily analytical. This is in response to the major theoretical contributions made by three commentators, Paul Krugman (1991), Michael Porter (1990) and Allen Scott (1988). Although their analytical approaches have differed widely, in that Scott (1988) and Porter (1998) use non-technical approaches and observation-lead approaches, whereas Krugman (1991) adopted a technical microeconomic approach, for all three commentators, the focus of analysis has been on the potential scale economy and efficiency benefits of industrial clustering. Until their work focused attention on the role played by geography, cities and clusters in the determination of national and international trade patterns and growth, most economists outside of the regional science tradition were largely unaware of the importance of these issues. Their work has subsequently opened up discussions of the role which clustering and agglomeration play in economics and business matters to a much wider academic and policy-making audience than was previously the case. In particular, following the insights of new growth theory (Romer, 1986, 1987; Lucas, 1988), the role played by local knowledge spillovers and local labour market skills is now seen to be essential in developing an understanding of local growth processes.

It can be demonstrated, however, that an unequal spatial distribution of activity is a natural outcome of a random process (Ellison & Glaeser, 1997), without any recourse to arguments about economies of scale (Gabaix, 1999a,b). It is therefore clear that other empirical evidence of the importance of locational characteristics is also required in order to provide a justification for the recent interest in cities and industrial clusters.

Fortunately, at the same time as the new theoretical agenda was being developed, there have also been two major empirical agendas developed which indicate the importance of cities and industrial clusters as sources of economic growth. The first is the agenda developed by Glaeser (Glaeser et al., 1992; Ellison & Glaeser, 1997; Gaspar & Glaeser, 1998) which has focused on the development of new ways to identify and evaluate agglomeration economies. Meanwhile, the second empirical agenda is developed on the basis of the work of Jaffe (Jaffe, 1989; Jaffe et al., 1993) and Acs and Audrestch (Audrestch & Feldman, 1996; Audretsch & Stephan, 1996; Acs, 2002), and focuses on ways of identifying empirically local knowledge spillovers using primarily patent and R&D data, and also in some cases spatial statistics (Anselin et al., 1997). These two approaches have produced a wealth of empirical evidence which suggests that economic growth processes are indeed localized in geographical space owing to the presence of agglomeration economies.

The second reason behind the increased interest in the relationship between locational characteristics and regional growth is a technological one. The primary technological development which has contributed to the renewed interest in the economic impacts of agglomeration has been the rapid improvement in information, communications and transportation technologies. These technological advances have improved the ability of corporate and government decision makers to coordinate either market or organizational activities across progressively larger geographical areas. Modern discussions about globalization (Michie, 2003) have highlighted the ability of corporate organizations to exploit resources in a more sophisticated manner across a wider geographical extent. Although the evidence on this point is rather mixed (McCann, 2004) with certain key exceptions, Glaeser (1998) argues that taking a broad view of all the empirical evidence indicates that the aggregate share of total output accounted for by transportation costs has fallen markedly over time (Glaeser & Kohlhasse, 2004). This has led to the greater geographical extending of many markets. At the same time, however, rather than reducing the importance of space and location, modern technology (Gaspar & Glaeser, 1998) is also argued to be partly responsible for the increasing dominance of particular service-oriented (Glaeser, 2005) and 'global' cities (Gordon, 2002; Buck et al., 2005) within the international financial system. This is apparently due to the increased quantity and complexity of the information generated and handled by information technology, and the resulting increased face-to-face interaction required in order to manage and act on this information. Technology has therefore given geographical proximity a renewed importance.

The third reason behind the increased interest in the relationship between locational characteristics and regional growth is an institutional one. Over the last two decades, there have been widespread institutional changes within the global and regional trade frameworks. The movements towards free-trade and integrated market areas such as the EU, NAFTA, ASEAN and MERCOSUR, have meant that the tariff structures associated with national borders may be becoming progressively less important in terms of their effects in shaping a nation's economic performance (Clement et al. 1999; Yeung et al., 1999). The role of trade barriers in determining the geographical patterns of trade is being progressively eroded while, at the same time, such institutional changes also provide for the ever-increasing two-way flows of both labour and capital across space. As such, explanations of the geography of trade patterns based on national factor proportions become progressively less realistic, while explanations based explicitly on the structure of economic geography, and in particular the urban system, become more important.

For these three analytical, technological and institutional reasons, agglomeration economics and the economics of industrial clustering have emerged over the last two decades as central issues of research. For both urban and regional economists and also economic geographers, the increased interest in these topics is obviously very welcome indeed. However, as the quantity and variety of research in this field has burgeoned, the mixing, and sometimes rather liberal use, of technical terminology, has led to a certain amount of analytical overlapping, confusion and duplication, some of which has significant problematic implications for the way we evaluate issues and design public policy. For this reason, it is necessary to consider the nature of agglomeration economics and its links with economic growth, the origins and analytical foundations of different views and hypotheses regarding the advantages of agglomeration, the structural assumptions underlying agglomeration and clustering.

3 Agglomeration, clusters and innovation

In order to account for the fact that economic activities and people are generally clustered together in space, it is necessary in part to adopt the notion that economies of scale can be place-specific. This assumption is essential, because the clustering of activities in space increases competition for land, and in turn increases in nominal local labour prices are required in order to maintain real wages. The overall resulting increases in both local land and labour prices will increase the costs and reduce the profitability of local firms, unless there are some more than compensating benefits associated with clustering. As has already been mentioned, it can be demonstrated that an unequal spatial distribution of activity is a natural outcome of a random process (Ellison & Glaeser, 1997; Gabaix, 1999a,b), without making any assumptions about the operation of economies of scale. However, in order for such uneven patterns not to disappear in response to localized factor price inflation, we still require some positive localized advantages to exist in order to sustain an equal expected growth rate in large as well as small clusterings of activity. Consequently, it is not necessary for us to assume systematically faster growth in larger cities in order to explain clustering and agglomeration, but it is necessary to accept that some clustering and agglomeration advantages must exist in order to explain why growth does not tend to slow down systematically as centres grow and rents rise.

Such location-specific economies of scale are generally known as agglomeration economies, and following Marshall's (1920) original discussion these economies are generally understood to be external economies, which are independent of a single firm, but which accrue to all of the firms located in the same area. Marshall provided three reasons why such localized economies of scale might exist, namely local knowledge spillovers, local non-traded inputs produced under scale economies, and a local skilled labour pool.

Firstly, industrial clustering allows for frequent direct informal face-to-face contact between individuals, and this may allow for tacit knowledge to be shared between firms. The advantage of spatial clustering in this case is therefore that proximity maximizes the mutual accessibility of all individuals within the cluster, thereby improving the knowledge spillovers available to all local participants. This may be particularly advantageous in market environments characterized by rapidly changing information. Secondly, industrial clustering provides for the possibility that certain specialist inputs can be provided to the local group in a more efficient manner than would be the case if all of the firms were geographically dispersed. Thirdly, the spatial grouping of firms also allows for the creation

of a local specialized labour pool, thereby reducing labour hiring and search costs, and also providing a risk reduction mechanism in the face of firm-specific demand fluctuations.

Directly observing these Marshallian effects in reality can be very difficult. Therefore, in practice, urban and regional economists generally employ an Ohlin–Hoover classification (Ohlin, 1933; Hoover, 1937, 1948), which splits agglomeration economies into three types, namely internal returns to scale, localization economies, and urbanization economies. Internal returns to scale do not strictly fit with the Marshallian notion of agglomeration, which is based on externalities. However, location-specific economies of scale which are internal to a firm imply that significant stocks of both labour and capital will be spatially concentrated at one particular location. This can then contribute to the development of the other two forms of agglomeration economies, namely localization economies and urbanization economies. The difference between these two cases is that, if the agglomeration benefits accrue to activities in the same sector located at the same place these are termed 'localization economies', whereas, if the benefits accrue to a diverse range of local sectors, they are termed 'urbanization economies'. In terms of the mechanisms by which agglomeration externalities are transmitted, each of Marshall's sources of agglomeration can contribute to the development of either localization economies or urbanization economies. However, while the distinction between the Ohlin–Hoover agglomeration categories initially appears to be primarily related to the definitional boundaries of the firms and sectors in which the externality benefits accrue locally, there are also qualitative differences in the observed outcomes. In the case of localization economies, the Marshallian externalities are generally associated with the development of industrial clusters of particular sectors. On the other hand, in the case of urbanization economies, the externality benefits accrue to firms across different sectors (Jacobs, 1960) located in the same place. It is these overall urbanization economies which are normally regarded as the typical agglomeration benefits associated with large cities (Duranton & Puga, 2000), whereby the city-urbanization effects are generally assumed to comprise a range of lower-order localization economies. On the other hand, in the case of smaller cities, it may be that a small number of individual sectors will tend to dominate the city, rather than the city dominating the individual sectors (Duranton & Puga, 2000).

The notion that cities and clusters represent location-specific economies of scale, however, is still a rather static notion which is hard to relate to economic growth, unless we also introduce some dynamic elements to the activities and characteristics of cities and clusters. Although the simple Marshallian description allows for knowledge spillovers, no discussion of the types of knowledge or the role of the knowledge spillovers in local economic growth is provided; this can be achieved by incorporating notions of knowledge acquisition, learning and the evolution in knowledge within the spillovers-type framework. While modern growth theory emphasizes the role played by knowledge in the aggregate macro-growth process, modern thinking about agglomeration and clusters emphasizes the intermediate role played by cities in facilitating the process of knowledge generation, information transmission and firm creation. The analytical focus of agglomeration economics is therefore on the relationship between industrial clusters, cities and innovation, and the study of agglomeration economics has consequently emerged as something of a hybrid subject, which combines insights regarding increasing returns to scale and externalities from economics, with insights regarding processes of innovation

from management studies, science and technology policy analysis, regional planning and economic geography.

The hybrid nature of the subject is also reflected in the variety of models and analytical approaches which are adopted. There are currently five families of models, plus variants thereof, which deal with the relationship between industrial clusters, agglomeration and innovation, and which are widely used in the literature. These are the growth pole model (Perroux, 1950; Boudeville, 1966), the incubator or nursery-cities model (Chinitz, 1961, 1964; Duranton & Puga, 2001), the product cycle model (Vernon, 1966, 1979; Markusen, 1985), the Porter (1990, 1998) clusters model and the new industrial areas and innovative milieux models (Scott, 1988; Saxenian, 1994; Aydalot & Keeble, 1988; Paniccia, 2002; Becattini, 2004). The first three families of models originally emerged during the 1960s, whereas the latter two families of models reflect developments during the last two decades. Not surprisingly, these various families of models tend to emphasize different aspects of the possible relationship between geography and innovation.

The growth pole model (Perroux, 1950) and its spatial manifestation (Boudeville, 1966) employed some of the ideas of Schumpeter (1934) in emphasizing the ways in which economic relationships can exhibit certain polarities regarding the particular network structures via which financial transactions are mediated. In particular, the decisions made by large 'polar' firms have major financial implications for the behaviour of other firms via the logic of buyer–supplier relationships. As such, many firms may not be independent or free to act of their own accord, and therefore the innovation behaviour of an individual firm alone may not be of significance. Rather, firm innovation behaviour may only be both possible when it is congruent with the behaviour of the key large firms linked in the same financial networks. On this argument, knowledge of these network structures and the behaviour of key large local firms will be essential for innovation, and proximity may engender such knowledge, thereby leading to significant local growth effects. At the other end of the analytical spectrum, however, is the incubator or nursery-city model of Chinitz (1961) and Duranton and Puga (2001), which argues that cities which are highly diversified and which contain a broad range of different types of industries and firm sizes will act as superior 'incubators' for the development and growth of both new and small firms, because in such environments there will be a variety of local business services available for supporting these small firms. On this argument, knowledge, the size distribution of a city or cluster is critical for innovation. Therefore, while the growth pole and the incubator hypotheses both suggest that geographical proximity is essential for innovation, the former hypothesis focuses on the key role played by large firms, whereas the latter hypothesis focuses on the importance of sectoral diversity.

The third alternative analytical approach to dealing with the relationship between industrial clusters, agglomeration and innovation is the product cycle model (Vernon, 1960; Markusen, 1985). The original product cycle argument suggested that the relationship between geography and innovation depended on the relationship between the location behaviour of firms and the structure of the urban system (Vernon, 1960). In particular, firms will tend to separate activities by location according to the stage in the life-cycle of the product. Following the subsequent work of Vernon (1966), later explicitly spatial versions of this model assume that the innovation behaviour of a geographical area is directly related to the way the location behaviour of large oligopolistic firms (Markusen, 1985) is related to the structure of the urban system. The general conclusions

of this model are that, in dominant central cities, such oligopolistic firms will tend to locate information-intensive activities such as R&D and high level decision making, all of which relate to the early stages of the product life-cycle, while at the same time locating facilities producing more mature, less novel and rather standardized products in more geographically peripheral areas with lower labour costs and skills. The outcome of this behaviour is that there will tend to emerge a clear separation and qualitative distinction between activity types in central city-regions and those in more peripheral areas. In particular, only central areas will be characterized by the production of innovative activities.

Although each of these three approaches to understanding the relationship between industrial clustering, cities and innovation first emerged during the 1960s, they underpin many current lines of analysis. In addition, however, the two more recent approaches to understanding the relationship between clusters, cities and innovation have been developed. These two new models are the clusters model of Porter (1990) and the new industrial spaces model (Scott, 1988).

Porter's (1990) cluster model argues that geographical proximity facilitates mutual visibility and transparency between competitors. In others words, firms are able to observe the competitive developments of each other, and this mutual visibility and transparency itself acts as a spur to all local firms to continue to improve their own individual competitiveness. The result of this process of localized competition is that the competitiveness of the local industrial cluster as a whole is increased. This argument applies equally to both large firms and small firms, although Porter assumes that spatial clustering of firms in the same sector is particularly important for small firms which rely mainly on external sources of information and technology.

The fifth and final approach to understanding the relationship between industrial clustering, cities and innovation is provided by the literature on new industrial areas (Scott, 1988). Recent observations and anecdotal evidence suggest that the growth of regions dominated by large numbers of small firms, such as Silicon Valley and the Emilia-Romagna region of Italy (Scott, 1988), appears to be related crucially to innovation, at least in certain industrial sectors. These observations have led to suggestions from many observers that industries which are made up of spatial clusters of small firms tend to be more highly innovative than industries comprising mainly large firms (Saxenian, 1994). The reason for this is that innovation is assumed to be most likely to occur in small and medium-sized enterprises, which have neither the scale nor the risk-bearing capacity to provide all of the key inputs on their own account, and the geographical proximity of SMEs is assumed to be a key criterion for the development of mutual trust relations based on a shared experience of interaction with decision-making agents in different firms. The origin of these arguments is in the social network theory analysis of environments which reduce opportunism and promote trust between local firms. These arguments regarding what is termed 'social capital' have recently been introduced into economics (Glaeser et al., 2000, 2002**)** and imply that certain types of social environments provide the appropriate 'milieux' for innovations to take place (Aydalot and Keeble, 1988), primarily because firms are willing to cooperate selectively as well as compete, and individuals are willing to share information. The key development here is the integration between human capital (as defined simply in terms of skills and education) and interpersonal networking to produce 'social capital'. Currently popular twists on these arguments have also been introduced by Florida (2003) who promotes the concept of 'creativity' on the part of

unorthodox 'bohemian' social groups, as a key driver of local, and particularly urban, economic development. The problem with the Florida concept of 'creativity', however, is that it is empirically indistinguishable from more general orthodox human capital arguments about the important role played by people who generate new ideas (Glaeser, 2004).

Agglomeration economies are to some extent, therefore, a 'black box', and each of these families of analytical models therefore provides a different insight and possible explanation of the workings of agglomeration economies. However, in order to identify the likely workings of a particular agglomeration or industrial cluster it is necessary to find a common system of reference, which allows us to compare and contrast these different approaches. By far the most parsimonious way to do this is to employ a transactions-costs framework in which the analytical focus is on the nature and stability of the relations between firms within an agglomeration. The original framework (Gordon & McCann, 2000) has since been used extensively in different contexts (Gordon & McCann, 2005a; McCann, 2004; McCann & Sheppard, 2003; McCann & Shefer, 2004; McCann & Mudambi, 2004, 2005; Simmie & Sennet, 1999) because it is both the simplest and also the most comprehensive organizing framework available for understanding the possible micro-foundations of industrial clusters and agglomerations.

4 Clusters, firm types and the nature of transactions

Adopting the transactions costs framework approach, we can see from the literature that there are three broad typologies of agglomerations and industrial clusters, defined in terms of the features they exhibit (Gordon & McCann, 2000). These are the *pure agglomeration*, the *industrial complex*, and the *social network*. The key feature which distinguishes each of these different ideal types of spatial industrial cluster is the nature of the relations between the firms within the cluster. The characteristics of each of the cluster types are listed in Table 2.1 and, as we see, the three ideal types of clusters are all quite different.

In the model of pure agglomeration, inter-firm relations are inherently transient. Firms are essentially atomistic, in the sense of having no market power, and they will continuously change their relations with other firms and customers in response to market arbitrage opportunities, thereby leading to intense local competition. Consequently, there is no loyalty between firms, nor are any particular relations long-term. The external benefits of clustering accrue to all local firms simply by reason of their local presence, the price of which is the local real estate market rent. There are no free-riders, access to the cluster is open, and consequently it is the growth in the local real estate rents which is the indicator of the cluster's performance. This idealized type is best represented by the pure Marshall (1920), Chinitz (1961) and Jacobs (1960) models, as well as the more recent new economic geography models (Fujita et al., 1999; Fujita & Thisse, 2002). The industrial structure represented by these models is that of monopolistic competition, and the notion of space is essentially urban space. In other words, this type of clustering only exists within individual cities. The empirical verification of this pure agglomeration phenomenon relies on evidence of localized productivity growth, associated with growth in local real estate prices, real wages and employment.

The industrial complex is characterised primarily by long-term stable and predictable relations between the firms in the cluster. This type of cluster is the type of spatial cluster typically discussed by classical (Weber, 1909) and neoclassical (Moses, 1958) location–production models, and represents a fusion of locational analysis with input–output

Table 2.1 Industrial clusters

Characteristics	Pure agglomeration	Industrial complex	Social network
Firm size	atomistic	some firms are large	variable
Characteristics of relations	non-identifiable fragmented unstable	identifiable stable trading	trust loyalty joint lobbying joint ventures non-opportunistic
Membership	open	closed	partially open
Access to cluster	rental payments location necessary	internal investment location necessary	history experience location necessary but not sufficient
Space outcomes	rent appreciation	no effect on rents	partial rental capitalization
Notion of space	urban	local but not urban	local but not urban
Example of cluster	competitive urban economy	steel or chemicals production complex	new industrial areas
Analytical approaches	models of pure agglomeration	location–production theory input–output analysis	social network theory (Granovetter)

analysis. Component firms within the spatial grouping each undertake significant long-term investments, particularly in terms of physical capital and local real estate, in order to become part of the grouping. Access to the group is therefore severely restricted by both high entry and exit costs, and the rationale for spatial clustering in these types of industries is that proximity is required primarily in order to minimize inter-firm transport transactions costs. Rental appreciation is not a feature of the cluster, because the land which has already been purchased by the firms is not for sale, nor are the informal knowledge spillovers evident in the model of pure agglomeration (McCann & Mudambi, 2004, 2005). This ideal type of cluster more closely reflects the internal returns to scale arguments of Hoover as well as aspects of the spatial growth pole model of Perroux (1950) and Boudeville (1966) and the spatial product cycle model of Markusen (1985). The literature on these types of inter-firm relations has typically focused on traditional commodities-based sectors (Isard & Kuenne, 1953) such as steel, chemical and pharmaceuticals. However, this notion of stable and well-defined inter-firm relations is equally applicable to any modern high-technology sectors running tightly-coordinated supply chains or Just-In-Time (JIT) production systems, such as aerospace, automobiles, electronics (McCann, 1997), or more standardized or lower-technology contemporary commercial sectors, such as food processing or supermarket retailing. Even some elements of the financial services industry such as retail banking and mortgage provision will fall into this grouping. The key issue here is that the industrial structure within the complex exhibits primarily oligopoly characteristics, and the notion of space in the complex model is local, but not necessarily urban. In other words, these types of local industrial clusters can exist either within, outside or beyond the limits of an individual city. The empirical verification of the industrial complex phenomenon relies on evidence of long-term and stable inter-firm

transactions (McCann, 1997) and good examples of the types of possible approaches here are demonstrated by Feser and Bergman (2000), Feser and Sweeney (2000, 2002) and Feser et al. (2005).

The third type of spatial industrial cluster is the social network model. This is associated primarily with the work of Granovetter (1973, 1985), and is a response to the markets and hierarchies model of Williamson (1975). Whereas the pure agglomeration model and the industrial complex described above represent the clustered spatial equivalents of the market and the hierarchy alternative modes of coordination, the social network model argues that mutual trust relations between key decision-making agents in different organizations may be at least as important as decision-making hierarchies within individual organizations. The key feature of such trust relations is an absence of opportunism, in that individual firms will not fear reprisals after any reorganization of inter-firm relations. Inter-firm cooperative relations may therefore differ significantly from the organizational boundaries associated with individual firms, and these relations may be continually reconstituted. All of these behavioural features rely on a common culture of mutual trust, the development of which depends largely on a shared history and experience of the decision-making agents. This social network model is essentially aspatial, but, from the point of view of geography, it can be argued that spatial proximity will tend to foster such trust relations, thereby leading to a local business environment of confidence, risk taking and cooperation. Spatial proximity is necessary but not sufficient to acquire access to the network. As such, membership of the network is only partially open, in that local rental payments will not guarantee access, although they will improve the chances of access. The social network model therefore contains some elements of the Porter model (1990, 1998), but primarily it is associated with the new industrial areas model (Scott, 1988). In this model space is once again local, but not necessarily urban, and may extend across a broader definition of a city-region. These trust relations will be manifested by a variety of features, such as joint lobbying, joint ventures, informal alliances and reciprocal arrangements regarding trading relationships.

In reality, all spatial concentrations of economic activity will exhibit at least one of these cluster types, and the case of large cities will exhibit aspects of all three models. An important point here is that pure agglomerations are most definitely clusters, albeit one particular and specific type of cluster. The approach of some economic geographers which assumes that only the social network model really represents a genuine cluster is analytically nonsense, and the comparison between the different cluster concepts provides powerful insights. The reason is that all three of these industrial cluster types exhibit economies of scale which can compensate either for local factor price appreciation or for the costs involved in the overcoming of geographical space.

Following these arguments, if we assume that one of the major functions of cities, agglomerations and industrial clusters is not only to reduce the costs of spatial inter-firm transactions but also to generate new ideas and new firms, then each of these three models of agglomeration clustering can also be associated with increased levels of innovation relative to other locations. This is very important when we consider macroeconomic growth as well as regional growth. From an empirical perspective, however, the important point is to identify exactly what are the dominant transactions costs features of the agglomeration or industrial cluster in question, by assessing the stability, longevity and loyalty of inter-firm transactions and relations. Only by doing this can we clearly identify the microeconomic rationale for the existence of each agglomeration or industrial cluster.

5 Agglomeration measurement problems

Measuring agglomeration effects directly is problematic for the reasons outlined in the above section. Pure agglomeration, by definition, is an externality phenomenon and, as such, cannot be observed directly. Therefore, such effects can only be observed indirectly by observing, for example, the growth in local real estate prices and the associated growth in the local real wages and employment, as these will be the only tangible outcomes of the pure form of agglomeration. The productivity effects of agglomeration appear to be related to employment density (Ciccone & Hall, 1996). However, the local density of employment activities is not of itself sufficient evidence of agglomeration because, as we have seen, the pure agglomeration is also characterized by frequent and unstable trading relations among broadly atomistic firms. Evidence for the prevalence of these types of trading relations must also be provided. On the other hand, the empirical verification of the industrial complex phenomenon is based on evidence of long-term and stable inter-firm transactions among oligopolistic firms, along with local employment and wage growth, while the types of evidence suggestive of the existence of local social networks will be joint lobbying, joint ventures, informal alliances and reciprocal arrangements regarding trading relationships, between primarily small firms, along with employment and wage growth. Once again, in each of these cases, evidence for the prevalence of the appropriate types of trading relations or alliances must also be provided.

Unfortunately, in reality, evidence of particular types of trading relations or alliances is often very hard to find, as it requires large-scale micro-level surveys (Arita & McCann, 2000; Gordon & McCann, 2000, 2005b), and the empirical identification of the particular scale effects which contribute to specific examples of industrial clustering is therefore very difficult. As such, the actual effect of agglomeration externalities on each industrial sector, or range of sectors, is in general a priori not identifiable, and therefore ex post indirect measures are normally adopted to try to identify the effects of such externalities. In practice, observations of sectoral employment distributions in particular localities have become the most commonly used method of identifying the operation of either urbanization or localization economies. In situations of industrial clustering and employment density, indices of sectoral employment specialization and diversity are employed as proxies for various types of local agglomeration effects. In particular, indices which point to local employment specialization are treated as evidence of localization economies, while indices which point to employment diversity are treated as evidence of urbanization economies. This approach, however, is not as straightforward as might be supposed. It is a standard technique in applied regional input–output modelling to relate local employment patterns to trade patterns (Dewhurst & McCann, 1998; McCann, 2001) and this is done by treating data on the former as a proxy for the latter. However, this is a very indirect approach, the weakness of which is that it is based on the assumption of universal Leontief production and consumption technologies. In a similar vein, using measures of local employment variety as evidence of either localization or urbanization economies is also problematic, because such an approach is based on the strong new economic geography assumption that there is a direct correspondence between the variety of local production, the variety of local employment and the variety of regional trade relationships.

There are four major problems associated with using local labour or employment data in the assessment of agglomeration externalities. Firstly, using employment data to distinguish between localization economies (Marshall, 1920) and urbanization economies

(Hoover, 1948; Jacobs, 1960) is problematic from an empirical perspective, because the results of these models appear to be sensitive to the level of sectoral aggregation used (Glaeser et al., 1992; Henderson et al., 1995). Unfortunately, there is no theoretical guidance as to what is the most appropriate level of sectoral disaggregation. Moreover, the distinction between the three Ohlin–Hoover agglomeration classifications is rather arbitrary in many cases, given that mergers and acquisitions mean that firms are frequently changing ownership and sectors without necessarily changing either their locations or the nature of many of their transactions. This latter issue is particularly pertinent in the case of many service industries (Cohen, 1998).

Secondly, identifying whether industrial clustering is indeed actually due primarily to the existence of localized externalities, or rather is simply the outcome of similar location optimization behaviour by similar firms (McCann, 1995), is also empirically very difficult, and really requires additional microeconomic data on transactions and buyer–supplier relationships (Gordon & McCann, 2000, 2005b).

Thirdly, the various measures and indices of industrial specialization, based on local sectoral employment distributions which are available to us, themselves often produce quite conflicting results, with the relative rankings of different places being rather unstable, depending on which index is employed (Dewhurst & McCann, 2002).

Fourthly, the results of diversity index calculations types are very sensitive to the size of the geographical areas employed for the data aggregation (Dewhurst & McCann, 2007). It is a well-known observation that larger cities are generally more industrially diversified than smaller cities (Duranton & Puga, 2000), and this tends to suggest that urbanization economies dominate in larger cities while localization economies tend to dominate in smaller cities. However, the smaller is the area of analysis, *ceteris paribus*, the greater will be the apparent levels of specialization, while the larger is the area, the greater will be the apparent level of diversity. This endogeneity of the area size is a variation on the modifiable unit area problem (Openshaw & Taylor, 1979), and needs to be controlled for.

Moreover, even if we are able to overcome these four measurement problems, in addition, there are two other problems with empirically confirming that knowledge spillovers are a driver of localized agglomeration effects, and this concerns the evidence regarding the link between innovation and cities.

Firstly, while there is much evidence to suggest that in many cases the critical distance over which urban agglomeration externalities operate may be that of the city-metropolitan area (Gordon & McCann, 2005b), as is assumed by the pure agglomeration model. However, there is also much evidence to suggest that, for many firm-types and industries, the critical distances over which urban agglomeration externalities operate may be very much larger than that of the city-region (Arita & McCann, 2000; Caniels, 2000; Cantwell & Iammarino, 2000, 2003; Simmie, 1998; Suarez-Villa & Walrod, 1997), as is assumed by the industrial complex model and the social network model. Spatial econometrics may be able to help identify such critical distances, but this would involve re-estimating spatial models of the same overall areas for different nested subsets of regions.

Secondly, there is also much evidence to suggest that, while the link between innovation and cities can be strong in certain sectors (Acs, 2002), the evidence on these issues is not always conclusive. Cities do not always appear to be centres of innovation, nor does innovation necessarily appear to be centred on cities (Simmie, 2001; Simmie & Sennet,

1999). The reasons for these many exceptions to the simple standard hypotheses depend on how both the origins of innovation, and also the processes by which innovation emerges, relate to the different cluster concepts outlined in Table 2.1. These are complicated issues and depend on the nature of both technical and technological change. A detailed discussion of these issues can be found in Iammarino and McCann (2006).

Each of the above issues implies that identifying and measuring agglomeration economies is actually a major analytical and empirical challenge. Moreover, the balance between local specialization and diversity appears to have no effect on real wage growth (Glaeser et al., 1992) or on labour productivity (Henderson, 2003; Henderson & Thisse, 2004). This evidence implies that indices of sectoral employment diversity may not be so powerful for distinguishing between localization or urbanization economies because, in actual fact, industrial sectors may not be so relevant in explaining agglomeration economies. This largely accords with the transactions costs argument in section 3, which suggests that firm size and the nature of transactions and inter-firm relations is critical for understanding the nature of agglomeration, rather than the sectoral distributions per se. All three cluster typologies are consistent with colocated firms being either from the same sector or from a variety of different industries, although the pure agglomeration is probably more likely to be sectorally diversified than the industrial complex or the social network.

A final point concerns the issue of declining clusters. Areas such as spatially grouped industries or cities, which previously exhibited both a geographic concentration of activities and also appreciations in factor returns, but which now experience declining returns and investment, pose both an empirical and an analytical challenge to our concepts of agglomeration and clustering. In terms of our definition of clusters, such declining areas still exhibit a relatively high geographical concentration of activities, in comparison to many other localities. In this sense they are cities and clusters. However, declining factor returns implies that the congestion aspects of clustering are not compensated for by location-specific positive externalities. In this sense, these are no longer either agglomerations or clusters. Rather, such localities can be classified simply as cities or as spatially grouped industries, the characteristics of which depend on previous historical issues rather than on current productivity performance (McCann, 1995), unless the previously depressed locality is now once again exhibiting a resurgence (McCann, 1995), as is the case with the modern gentrification of many large cities which had initially emerged as industrial centres in the nineteenth century (Boddy & Parkinson 2004).

6 Conclusions

This chapter has reviewed the reasons for the dramatic increase in interest over recent decades in research concerning agglomeration economics and industrial clustering. As we have seen, this interest has arisen from a variety of sources, from a variety of analytical approaches and disciplines, and from a variety of observations from different contexts. While this burgeoning interest is very welcome indeed, it is important to consider the analytical origins and foundations of these different approaches and their associated terminology, because many of these approaches are based on different implicit assumptions. A comparison of these different analytical logics allows us to identify the specific ways in which we might be able to identify both the micro-working of particular types of agglomeration economies, and also empirical challenges that we need to overcome.

References

Acs, Z. (2002), *Innovation and the Growth of Cities*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
Anselin, L., A. Varga and Z. Acs (1997), 'Local geographical spillovers between university research and high technology innovations', *Journal of Urban Economics*, **42**, 422–48.
Arita, T. and P. McCann (2000), 'Industrial alliances and firm location behaviour: some evidence from the US semiconductor industry', *Applied Economics*, **32**, 1391–403.
Audretsch, D.B. and M.P. Feldman (1996), 'R&D spillovers and the geography of innovation and production', *American Economic Review*, **86**(3), 630–40.
Audretsch, D.B. and P.E. Stephan (1996), 'Company–scientist locational links: the case of biotechnology', *American Economic Review*, **86**, 641–52.
Aydalot, P. and D. Keeble (1988), *Milieux Innovateurs en Europe*, Paris: GREMI.
Barro, R.J. and X. Sala-i-Martin (1992), 'Convergence', *Journal of Political Economy*, **100**, 223–51.
Becattini, G. (2004), *Industrial Districts: A New Approach to Industrial Change*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
Boddy, M. and M. Parkinson (2004), *City Matters: Competitiveness, Cohesion and Governance*, Bristol: Policy Press.
Borts, G.H. and J.L. Stein (1964), *Economic Growth in a Free Market*, New York: Free Press.
Boudeville, J.R. (1966), *Problems of Regional Planning*, Edinburgh: Edinburgh University Press.
Buck, N., I.R. Gordon, A. Harding and I. Turok (eds) (2005), *Changing Cities*, London: Routledge.
Caniels, M.C.J. (2000), *Knowledge Spillovers and Economic Growth*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
Cantwell, J.A. and S. Iammarino (2000), 'Multinational corporations and the location of technological innovation in the UK regions', *Regional Studies*, **34**(4), 317–32.
Cantwell, J.A. and S. Iammarino (2003), *Multinational Corporations and European Regional Systems of Innovation*, London: Routledge.
Chinitz, B. (1961), 'Contrasts in agglomeration: New York and Pittsburgh', *American Economic Review*, **51**, 279–89.
Chinitz, B. (1964), *City and Suburb*, Englewood Cliffs, NJ: Prentice-Hall.
Ciccone, A. and R.E. Hall (1996), 'Productivity and the density of Economic activity', *American Economic Review*, **86**, 54–70.
Clement. N.C., G. del Castillo Vera, J. Gerber, W.A. Kerr, A.J. MacFadyen, S. Shedd, E. Zepeda and D. Alarcon (1999), *North American Economic Integration: Theory and Practice*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
Cohen, B. (1998), *The Geography of Money*, Ithaca, NY: Cornell University Press.
Dewhurst, J.H.L. and P. McCann (1998), 'Regional size, industrial location and input–output coefficients', *Regional Studies*, **32**(5), 435–44.
Dewhurst, J.H.L. and P. McCann (2002), 'A comparison of measures of industrial specialization for travel-to-work areas in Great Britain, 1981–1997', *Regional Studies*, **316**(5), 541–51.
Dewhurst, J.H.L. and P. McCann (2007), 'Specialisation and regional size', in B. Fingleton (ed.), *New Directions in Economic Geography*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
Dixon, R.J. and A.P. Thirlwall (1975), *Regional Growth and Unemployment in the United Kingdom*, London: Macmillan.
Duranton, G. and D. Puga (2000), 'Diversity and specialisation in cities: where and when does it matter?', *Urban Studies*, **37**(3), 533–55.
Duranton, G. and D. Puga (2001), 'Nursery cities: urban diversity, process innovation, and the life cycle of products', *American Economic Review*, **91**(5), 1454–77.
Ellison, G. and E.L. Glaeser (1997), 'Geographic concentration in US manufacturing industries: a dartboard approach', *Journal of Political Economy*, **105**, 889–927.
Feser, E.J. and E.M. Bergman (2000), 'National industry cluster templates: a framework for applied regional cluster analysis', *Regional Studies*, **34**, 1–19.
Feser, E.J. and S.H. Sweeney (2000), 'A test for the coincident economic and spatial clustering of business enterprises', *Journal of Geographical Systems*, **2**, 349–73.
Feser, E.J. and S.H. Sweeney (2002), 'Theory, methods, and a cross-metropolitan comparison of business clustering', in P. McCann (ed.), *Industrial Location Economics*, Cheltenham, UK and Northampton, MA, US: Edward Elgar, pp. 222–59.
Feser, E.J., S.H. Sweeney and H. Renski (2005), 'A descriptive analysis of discrete US industrial complexes', *Journal of Regional Science*, **45**(2), 395–415.
Florida, R. (2003), *The Rise of the Creative Class*, New York: Basic Books.
Fujita, M. and J.F. Thisse (2002), *Economics of Agglomeration*, Cambridge: Cambridge University Press.
Fujita, M., P. Krugman and A.J. Venables (1999), *The Spatial Economy: Cities, Regions and International Trade*, Cambridge, MA: MIT Press.

Gabaix, Z. (1999a), 'Zipf's law and the growth of cities', *American Economic Review: Papers and Proceedings*, **89**(2), 129–32.
Gabaix, Z. (1999b), 'Zipf's law for cities: an explanation', *Quarterly Journal of Economics*, **114**(3), 739–67.
Gaspar, J. and E.L. Glaeser (1998), 'Information technology and the future of cities', *Journal of Urban Economics*, **43**, 136–56.
Glaeser, E.L. (1998), 'Are cities dying?', *Journal of Economic Perspectives*, **12**(2), 139–60.
Glaeser, E.L. (2004), 'Book review of Richard Florida's *The Rise of the Creative Class*' (http://post.economics.harvard.edu/faculty/glaeser/papers.html).
Glaeser, E.L. (2005), 'Urban colossus: why is New York America's largest city?', *Federal Reserve Bank of New York Economic Policy Review*, **11**(2), 7–24.
Glaeser, E.L. and J. Kohlhasse (2004), 'Cities, regions and the decline of transport costs', *Papers in Regional Science*, **83**(1), 197–228.
Glaeser, E.L., D.I. Laibson and B. Sacerdote (2002), 'An economic approach to social capital', *Economic Journal*, **112**(483), 437–58.
Glaeser, E.L., D.I. Laibson, J. Scheinkman and C.L. Soutter (2000), 'Measuring trust', *Quarterly Journal of Economics*, **115**, 811–46.
Glaeser, E., H.D. Kallal, J.A. Schinkmann and A. Shleifer (1992), 'Growth in cities', *Journal of Political Economy*, **100**, 1126–52.
Gordon, I.R. (2002), 'Global cities, internationalization and urban systems', in P. McCann (ed.), *Industrial Location Economics*, Cheltenham, UK and Northampton, MA, US: Edward Elgar, pp. 187–206.
Gordon, I.R. and P. McCann (2000), 'Industrial clusters, complexes, agglomeration and/or social networks?', *Urban Studies*, **37**, 513–32.
Gordon, I.R. and P. McCann (2005a), 'Clusters, innovation and regional development: an analysis of current theories and evidence', in B. Johansson, C. Karlsson and R. Stough (eds), *Entrepreneurship, Spatial Industrial Clusters and Inter-Firm Networks*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
Gordon, I.R. and P. McCann (2005b), 'Innovation, agglomeration and regional development', *Journal of Economic Geography*, **5**(5), 523–43.
Granovetter, M. (1973), 'The strength of weak ties', *American Journal of Sociology*, **78**, 1360–89.
Granovetter, M. (1985), 'Economic action and social structure: the problem of embeddedness', *American Journal of Sociology*, **91**, 481–510.
Henderson, J.V. (2003), 'The urbanization process and economic growth: the so-what question', *Journal of Economic Growth*, **8**(1), 47–71.
Henderson, J.V. and J.F. Thisse (eds) (2004), *Handbook of Regional and Urban Economics*, vol. 4, Amsterdam: North Holland.
Henderson, J.V., A. Kuncoro and M. Turner (1995), 'Industrial development in cities', *Journal of Political Economy*, **103**, 1067–85.
Hoover, E.M. (1937), *Location Theory and the Shoe and Leather Industries*, Cambridge, MA: Harvard University Press.
Hoover, E.M. (1948), *The Location of Economic Activity*, New York: McGraw-Hill.
Iammarino, S. and P. McCann (2006), 'The structure and evolution of industrial clusters: transactions, technology and knowledge spillovers', *Research Policy*, **35**, 1018–36.
Isard, W. and R.E. Kuenne (1953), 'The impact of steel upon the greater New York–Philadelphia industrial region', *Review of Economics and Statistics*, **35**, 289–301.
Jacobs, J. (1960), *The Economy of Cities*, New York: Random House.
Jaffe, A.B. (1989), 'Real effects of academic research', *American Economic Review*, **79**, 957–70.
Jaffe, A.B., M. Trajtenberg and R. Henderson (1993), 'Geographic localization of knowledge spillovers as evidenced by patent citations', *Quarterly Journal of Economics*, **108**, 577–98.
Krugman, P. (1991), *Geography and Trade*, Cambridge, MA: MIT Press.
Lucas, R.E. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics*, **22**, 3–42.
Markusen, A. (1985), *Profit Cycles, Oligopoly and Regional Development*, Cambridge, MA: MIT Press.
Marshall, A. (1920), *Principles of Economics*, 8th edn, London: Macmillan.
McCann, P. (1995), 'Rethinking the economics of location and agglomeration', *Urban Studies*, **32**(3), 563–77.
McCann, P. (1997), 'How deeply embedded is Silicon Glen? A cautionary note', *Regional Studies*, **31**(7), 695–703.
McCann, P. (2001), *Urban and Regional Economics*, Oxford: Oxford University Press.
McCann, P. (2004), 'Urban scale economies: statics and dynamics', in R. Capello and P. Nijkamp (eds), *Urban Dynamics and Growth: Advances in Urban Economics*, Amsterdam: Elsevier, pp. 31–56.
McCann, P. and R. Mudambi (2004), 'The location decision of the multinational enterprise: some theoretical and empirical issues', *Growth and Change*, **35**(4), 491–524.
McCann, P. and R. Mudambi (2005), 'Analytical Differences in the economics of geography: the case of the multinational firm', *Environment and Planning A*, **37**(10), 1857–76.

McCann, P. and D. Shefer (2004), 'Location, agglomeration and infrastructure', *Papers in Regional Science*, **83**(1), 177–96.
McCann, P. and S.C. Sheppard (2003), 'The rise, fall and rise again of industrial location theory', *Regional Studies*, **37**(6–7), 649–63.
McCombie, J.S.L. and A.J. Thirlwall (1994), *Economic Growth and the Balance of Payments Constraint*, London: St. Martin's Press.
Michie, J. (2003), *The Handbook of Globalisation*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
Moses, L. (1958), 'Location and the theory of production', *Quarterly Journal of Economics*, **78**, 259–72.
Ohlin, B. (1933), *Interregional and International Trade*, Cambridge, MA: Harvard University Press.
Openshaw, S. and P.J. Taylor (1979), 'A million or so correlation coefficients: three experiments on the modifiable area unit problem', in N. Wrigley (ed.), *Statistical Applications in the Spatial Sciences*, London: Pion, pp. 127–44.
Paniccia, I. (2002), *Industrial Districts: Evolution and Competitiveness in Italian Firms*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.
Perroux, F. (1950), 'Economic space, theory and applications', *Quarterly Journal of Economics*, **64**, 89–104.
Porter, M.E. (1990), *The Competitive Advantage of Nations*, New York: Free Press.
Porter, M.E. (1998), 'Clusters and the new economics of competition', *Harvard Business Review*, **76**(6), 77–90.
Romer, P. (1986), 'Increasing returns and long-run growth', *Journal of Political Economy*, **94**, 1002–37.
Romer, P. (1987), 'Growth based on increasing specialization', *American Economic Review*, **77**(2), 56–72.
Saxenian, A. (1994), *Regional Advantage*, Cambridge, MA: Harvard University Press.
Schumpeter, J.A. (1934), *The Theory of Economic Development*, Cambridge, MA: Harvard University Press.
Scott, A.J. (1988), *New Industrial Spaces*, London: Pion.
Simmie, J. (1998), 'Reasons for the development of "Islands of innovation": evidence from Hertfordshire', *Urban Studies*, **35**(8), 1261–89.
Simmie, J. (ed.) (2001), *Innovative Cities*, London: Spon Press.
Simmie, J. and J. Sennett (1999), 'Innovative clusters: global or local linkages?', *National Institute Economic Review*, **170**, 87–98, October.
Solow, R.M. (1956), 'A contribution to the theory of economic growth', *Quarterly Journal of Economics*, **70**, 65–94.
Suarez-Villa, L. and W. Walrod (1997), 'Operational strategy, R&D and intrametropolitan clustering in a polycentric structure: the advanced electronics industries of the Los Angeles basis', *Urban Studies*, **34**(9), 1343–80.
Swan, T. (1956), 'Economic growth and capital accumulation', *Economic Record*, **32**, 334–61.
Vernon, R. (1960), *Metropolis 1985*, Cambridge, MA: Harvard University Press.
Vernon, R. (1966), 'International investment and international trade in the product cycle', *Quarterly Journal of Economics*, **80**(2), 190–207.
Vernon, R. (1979), 'The product cycle hypothesis in a new international environment', *Oxford Bulletin of Economics and Statistics*, **41**, 255–67.
Weber, A. (1909), *Über den Standort der Industrien*, translated by C.J. Friedrich (1929), *Alfred Weber's Theory of the Location of Industries*, Chicago, IL: University of Chicago Press.
Williamson, O.E. (1975), *Markets and Hierarchies*, New York: Free Press.
Yeung, M., N. Perdikis and W. Kerr (1999), *Regional Trading Blocs in the Global Economy: The EU and Asean*, Cheltenham, UK and Northampton, MA, US: Edward Elgar.