Urban Studies

Diversity and Specialisation in Cities: Why, Where and When Does it Matter?

Gilles Duranton and Diego Puga

The online version of this article can be found at:

Published by:

$SAGE

On behalf of:

Additional services and information for *Urban Studies* can be found at:

Email Alerts: http://usj.sagepub.com/cgi/alerts

Subscriptions: http://usj.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://usj.sagepub.com/content/37/3/533.refs.html

Diversity and Specialisation in Cities: Why, Where and When Does it Matter?

Gilles Duranton and Diego Puga

Summary. Why are some cities specialised and others diversified? What are the advantages and disadvantages of urban specialisation and diversity? To what extent do the structure of cities and the activities of firms and people in them change over time? How does the sectoral composition of cities influence their evolution? To answer these and related questions, we first distil some key stylised facts from the empirical literature on cities and the composition of their activities. We then turn to a review of different theories looking at such issues, and study the extent to which these theories contribute to the understanding of the empirical regularities.

1. Introduction

At a time when most central governments, at least in developed economies, have abandoned the idea of a strongly interventionist industrial policy, a new economic activism seems to be on the rise at the local level. Local authorities frequently strive for specific patterns of economic activities in their jurisdictions. Specific models of industrial organisations are explicitly targeted, with the industrial districts of the Third Italy or Silicon Valley as prime models to be followed (Bartik, 1996, gives a broad review of these issues). While not all of these interventions are necessarily misguided, many seem to lack a clear rationale or even to be based on common misconceptions.[1]

The issues relating to the composition of cities are necessarily complex. Why are some cities specialised and others diversified?

What are the advantages and disadvantages of urban specialisation and diversity? To what extent do the structure of cities and the activities of firms and people within them change over time? How does the sectoral composition of cities influence their evolution? To answer these and related questions, this paper starts by distilling some key stylised facts from the empirical literature on cities and the composition of their activities. We then turn to a review of different theories concerned with such issues, and study the extent to which these contribute to our understanding of empirical regularities. We structure these theories into three groups. First, those that are static in nature and focus on the geography within cities rather than on their spatial location with respect to each other. Secondly, static theories that look

Gilles Duranton is in the Department of Geography and Environment, London School of Economics, Houghton Street, London WC2A 2AE, UK. Fax: 0171 955 7412. E-mail: G.Duranton@lse.ac.uk. Diego Puga is in the Department of Economics, University of Toronto, 150 St George Street, Toronto, Ontario M5S 3G7, Canada. Fax: 416 946 5312. E-mail: dpuga@chass.utoronto.ca. The authors are very grateful to Duncan Black for giving them access to data on US cities, as well as for his comments. They also greatly benefited from comments by Henry Overman and Andrés Rodrõguez-Pose. The paper was partly written when the first author was visiting CESURE, Université de Lille 1, and while the second author was visiting the Institut d'Anàlisi Econòmica, Barcelona. The hospitality of both institutions is gratefully acknowledged.

more carefully at the spatial relations across cities. And, finally, theories that incorporate an explicit intertemporal dimension in the study of urban systems. A final section gathers some additional thoughts.

While concerned with policy issues, this survey does not aim to offer easy prescriptions. Instead, it highlights and discusses some of the complexities involved. But an additional motivation is present. There is a large body of extremely interesting empirical work on the structure of urban systems, but this often lacks an adequate theoretical base that could help to structure it. At the same time, related theoretical developments are not regularly subject to the discipline of assessing their correspondence with known empirical regularities. And both the empirical and theoretical research on the structure of cities is rarely well known to those not working in the field. Our main objective in this exercise of gathering, structuring and relating different empirical and theoretical approaches is to help build a common understanding of the nature and structure of cities.

2. Some Facts about Diversity and Specialisation in Cities

A large recent empirical literature has established a number of stylised facts about the composition of cities, their sizes and locations. In this context, note that gathering data on economic activity in cities is difficult. Since most of the recent research comes from the US, our exposition of the facts in this section is thus biased towards that country. Whenever possible, evidence about other countries is also given. US data and available data from the rest of the world often tend to go in the same direction as far as the issues examined here are concerned. However, it should be the task of future research to go into more depth and to corroborate or invalidate US findings. Another difficulty in exploring patterns of activities in cities involves the problems of definition raised by the concept of cities. In order to overcome these, to circumvent data limitations and to

limit the scope of this paper, we shall focus mostly on cities with more than 50 000 inhabitants. This is not neutral with respect to the theories that we examine in sections 3–5, since our stylised facts in this section ignore small cities and concentrate on medium-sized and large cities, which may be of a nature different from smaller cities. In this section, we summarise and structure facts about the composition of economic activity in cities under five headings.

Stylised fact 1. Specialised and diversified cities co-exist.

In nearly all countries, cities largely specialised in a few activities co-exist with much more diversified cities. To start looking at diversity and specialisation in cities in some detail, one needs to specify measures of specialisation and diversity. The simplest way to measure a city's specialisation in a given sector is to quantify the share of this sector in local employment. Since different cities are specialised in different sectors, one useful measure of specialisation for comparisons across cities is the employment share of each city's largest sector. If we denote s_{ij} as the share of industry j in city i, we can define the following specialisation index:

$$ZI_i = \underset{j}{Max}(s_{ij}) \qquad (1)$$

Of course, certain sectors account for a larger share of overall (say, national) employment than others. To correct for this, it is helpful to look at a city's relative (rather than absolute) specialisation, which can be measured as above, but dividing the share of each sector in local employment by its share in national employment. This leads naturally to the following relative-specialisation index:

$$RZI_i = \underset{j}{Max}(s_{ij}/s_j) \qquad (2)$$

where, s_j is the share of industry j in national employment.

Table 1 lists, for selected US cities, the local relative-specialisation index for two-digit manufacturing sectors in 1992 (with the corresponding sector in parentheses).[2] Black and Henderson (1998) provide a detailed description of the data set. Among the most

Table 1. Most and least specialised and diversified US cities in 1992

Rank	Specialisation		Diversity	
	City (sector)	RZI	City	RDI
1	Richmond, VA (tobacco)	64.4	Cincinnati, OH	166.6
2	Macon, GA (tobacco)	55.0	Oakland, CA	161.2
3	Lewiston, ME (leather)	49.6	Atlanta, GA	159.4
4	Galveston, TX (petroleum)	49.1	Philadelphia, PA	151.4
5	Bangor, ME (leather)	45.6	Salt Lake City, UT	120.8
6	Owensboro, KY (tobacco)	44.4	Buffalo, NY	110.1
7	Corpus Christi, TX (petroleum)	37.6	Columbus, OH	108.3
8	Cheyenne, WY (petroleum)	33.4	Portland, OR	94.1
315	Buffalo, NY (rubber and plastics)	1.6	Lawton, OK	2.4
316	Cincinnati, OH (chemicals)	1.5	Richland, WA	2.4
317	Chicago, IL (metal products)	1.5	Steubenville, OH	2.4

Source: Black and Henderson data set.

specialised cities, there are those specialised in sectors largely dependent on natural resources such as tobacco (Richmond, Macon, Owensboro), petroleum products (Galveston, Corpus Christi, Cheyenne) and leather (Laredo, Bangor, Lewiston). These are extreme cases for which the relative-specialisation index is often above 30, indicating the presence of a sector with a share in local employment 30 times larger than its share in national employment. We also find cities highly specialised in manufacturing sectors, which are less dependent upon local resources, such as textiles (Burlington, Danville, Anderson), industrial machinery and equipment (Rochester) or transport equipment (Flint). At the other extreme, we find cities with a very low relative-specialisation index (Cincinnati, Buffalo, Chicago), with no sector having a share of manufacturing employment larger than 1.6 times its national share.

Turning to diversity, a common measure is the inverse of a Hirshman–Herfindahl index. For each city i this sums over all sectors the square of each sector's share in local employment. Formally, the inverse Hirshman–Herfindahl index or diversity index is given by

$$DI_i = 1/\sum_j s_{ij}^2 \qquad (3)$$

If economic activity in the city under consideration is fully concentrated in a sector, we find $DI_i = 1$, and this index increases as activities in the city become more diverse.

Again, it is important to correct this measure for differences in sectoral employment shares at the national level. One possible way of making such a correction is to compute a relative-diversity index by summing for each city, over all sectors, the absolute value of the difference between each sector's share in local employment and its share in national employment. Formally, this leads to

$$RDI_i = 1/\sum_j |s_{ij} - s_j| \qquad (4)$$

This index increases the more the composition of activities in city i mirrors the diversity of the national economy. Table 1 lists, for selected US cities, the relative-diversity index based on two-digit manufacturing sectors in 1992. Note that, when measured in this way, diversity and specialisation are not exact opposites. A city with a main industry and a broad base of other industries can be both diversified and specialised. Nevertheless, the most diverse cities (Atlanta, Philadelphia and Salt Lake City) are also not very specialised in any particular sector.

One interpretation of the co-existence of diverse and specialised cities is that patterns of specialisation and diversity are purely

random. However, in their careful study of the agglomeration of industries in the US, Ellison and Glaeser (1997) show that this is highly unlikely: the distribution of economic activity in the US is too concentrated geographically at the level of most four-digit sectors for this to be the result of a random allocation. A second interpretation is that patterns of specialisation and diversification in cities merely mirror the spatial distribution of resources. For instance, cities in the French Alps or on the French Riviera tend to specialise in recreational activities for obvious topological and climatic reasons. This second possible explanation is more difficult to dismiss. Nevertheless, using panel-data econometrics, Henderson (1997a) is able to separate externalities from local fixed effects (i.e. persistent comparative advantage). He shows that even when fixed effects are accounted for, externalities are important. Thus, specialisation in cities is partly the result of economic interactions within a given sector (often called localisation economies), whereas diversity in cities is fostered by economic interactions across sectors (urbanisation economies).

Stylised fact 2. Larger cities tend to be more diversified. Cities of similar specialisation are of similar size.

We now turn to the relationship between city size and specialisation/diversity. It is widely believed that larger cities are more diversified. This assertion, while justified, needs some qualifications. First, there is a positive correlation between city size and the relative-diversity index described above, calculated on the basis of all private employment at the two-digit level in 1992, as shown in the plot of Figure 1. Note, however, that this link is not especially strong, partly because all cities have a large component of employment in non-tradeable activities. However, there is more to it than that. There are cases of specialised large cities, such as Los Angeles, which is a fairly specialised city (in entertainment) relative to its size. At this level of sectoral disaggregation, New York is also quite specialised in business services, although this sector comprises very heterogeneous activities. At the same time, smaller cities such as Buffalo, Columbus and Portland are very diversified.

Henderson (1997b) goes into more depth on this issue of diversity and size, and distinguishes between medium-sized cities
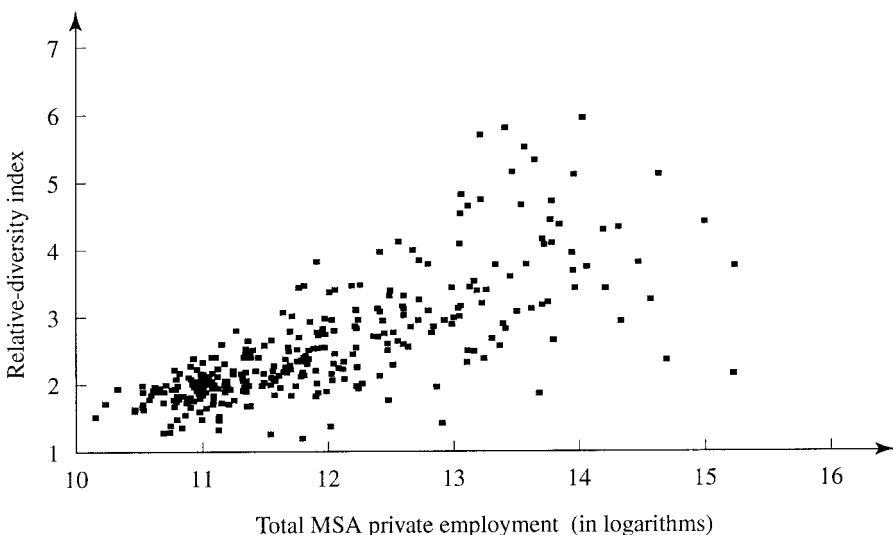
Figure 1. The size–diversity relation for US cities in 1992. *Source:* Black and Henderson (1998) data set.

(50 000–500 000 population) and large cities (over 500 000). He shows that large cities are, on average, more specialised in services (finance, insurance and real estate sectors) and less in manufacturing, than medium-sized cities. Looking at the industrial detail, medium-sized cities are more specialised in mature industries (such as textiles, food, pulp and paper) and less in new industries (such as electronic components and instruments). The contrast between instruments and textiles, which are sectors of comparable size, is particularly striking. In 1987, none of the 45 largest US metropolitan areas had more than 1 per cent of its workforce working in textiles, whereas 19 cities with a population of over 500 000 had between 1 and 5 per cent of their workforce in the instrument sector. Focusing on the size of specialised cities and distinguishing types of specialisation, Henderson (1988) shows that cities specialised in similar activities also tended to be of similar size. However, the evidence on this point is not very strong, given that standard deviations are often fairly high. Black and Henderson (1998), in their more comprehensive investigation of US cities for 1900–90, show that cities can be classified into groups of similar size/specialisation.

Stylised fact 3. The distribution of relative city sizes, individual city-size rankings and individual city specialisations are stable over time.

An examination of the distribution of city size in various countries reveals that this is very stable. This is all the more remarkable, given the growth of the average size of cities. However, as always in the analysis of the dynamics of distributions, looking at the shape of the distribution is not enough.[3] One needs to examine movements within this distribution—the ups and downs in the size of individual cities. Over the past 30 years in the US there are contrasting examples. For example, the size of the City of Phoenix increased by a factor of nearly 10 between 1950 and 1990, while the cities of Detroit and Pittsburgh both lost nearly half their population over the same period.[4] But the

systematic analysis by Black and Henderson (1998) of all US cities between 1900 and 1990 shows that Phoenix, Detroit and Pittsburgh are the exception, not the rule; in general, the relative sizes of US cities changed very little. Further, the current distribution of city size is very close to the ergodic (limiting) distribution, which we can interpret as evidence that the US urban system might be close to a 'steady state'. The general pattern is thus one of parallel urban growth, even though some cities gain more population than others. According to Eaton and Eckstein (1997), this pattern of overall stability is even stronger in France and Japan.

Not only are relative city sizes very stable, but there is also strong persistence of the same activities in the same cities. As with city size, there are some historical exceptions involving major overhauls at the sector level. These include the flight of textiles during the 18th century from the cities of Leiden and Haarlem in Holland to Manchester and Sheffield in the UK. Brezis and Krugman (1997) discuss these examples in further detail. More recent examples would include the successful economic restructuring of certain European cities formerly specialised in declining traditional activities, such as Cholet in France. The general pattern is, nonetheless, one of stability. Kim (1995) shows that the correlation of the coefficient of regional localisation for two-digit industries in the US between 1860 and 1987 at the state level is 0.64, which is surprisingly high, given the length of the period. Dumais *et al.* (1997) provide further evidence, showing that for US industries at the three-digit level, patterns of agglomeration have been rather stable over time. They show that industrial concentration in the US went up until 1947 and then down slightly until 1992 (which is consistent with the relative growth of larger cities, which are less heavily specialised). For most industries, the stability in their agglomeration patterns over 1972–92 is striking. The same picture of relative stability emerges from Henderson (1998) who studied in great detail 9 3-digit industries (4 high-tech industries

and 5 machinery industries) over 30 years. Even though, in the leading city, 7 out of 9 industries has changed between 1963 and 1992, specialisation patterns have tended to persist over time in these 9 industries.[5] Henderson (1997b) gives broader evidence of the stability of specialisation patterns for medium-sized cities in the US.

> *Stylised fact 4.* Individual city growth is related to specialisation and diversity, and to relative location.

Despite this overall picture of stability, some cities prosper, whereas others decay. A large part of the recent empirical literature has focused on trying to understand what characteristics of cities may contribute to their growth or decline. Glaeser *et al.* (1992) tackle this issue by asking whether diversity spurs further growth. Looking at employment growth in US cities, they find that diversity and local competition foster urban employment growth, whereas specialisation reduces urban employment growth. However, this finding could in part be a reflection of the recent relative decline in traditional manufacturing employment in the US, which tends to be more geographically concentrated (Black and Henderson, 1998). Some US and European cities specialised in traditional manufacturing activities were hit very hard by the decline of such activities. This suggests that distinguishing among sectors may be important. In fact, Henderson *et al.* (1995) find that, while urban diversity is, indeed, important for attracting new and innovative sectors, a history of similar past specialisation appears to matter more for mature industries. Furthermore, production of less-standardised or non-traditional items tends to be more concentrated in diversified metropolitan areas. On the other hand, production of traditional standardised items tends to be more heavily concentrated in cities which are often quite specialised in terms of the goods that are exported to other cities (Henderson, 1997a).

Urban growth also seems to be related to the relative locations of cities. Harris-Dobkins and Ioannides (1998), using US data

beginning in 1900, find that larger cities tend to grow at a slower pace. Larger cities also favour the development of adjacent cities as well of those located far away, but hinder the development of cities located at a medium distance (80–500 km). Black and Henderson (1998), using a different data set on US cities, find that the effect of lagged own-city size on growth is negative, but they find that the external market potential has a positive effect.[6] However, the positive effect of external market potential disappears if lagged own-city size is not considered (Ioannides and Overman, 1999). One possible interpretation of this is that cities tend to grow more when their market potential is large relative to their size, and that growth slows down as city size catches up. In any case, all these results are relatively weak.

While stylised facts about the effects on urban growth of relative location do not directly say anything about specialisation and diversity, these are nonetheless useful for discriminating between theories that do consider specialisation and diversity, but have different predictions regarding the link between urban growth and relative location. However, it must be said that the evidence to date remains very patchy and more work is needed on these issues.

> *Stylised fact 5.* There is a high rate of plant turnover. Most innovations take place in particularly diversified cities and most new plants are created there. Most relocations are from diversified to specialised cities.

The strong persistence of individual city specialisation discussed above is all the more surprising since data from several countries show that about 1 in 10 establishments closes down every year, to be replaced by new establishments (Dunne *et al.*, 1989a, 1989b; Davis and Haltiwanger, 1992). According to Dumais *et al.* (1997) more than one-half of all US manufacturing employees in 1992 worked in firms that did not exist in 1972, and nearly three-quarters of the plants existing in 1972 were closed by 1992. Even for surviving plants, changes have been dramatic. Among all plants still in operation in

1992 which had between 950 and 1050 employees in 1972, the mean for 1992 was 793 and the standard deviation 678. This would not be too surprising if new plants just replaced older plants, but Dumais *et al.* (1997) show that the birth process for new plants and the closure process are spatially very biased. The birth of new plants (in particular for new firms) tends to reduce the degree of agglomeration of particular sectors, suggesting that new plants are created in locations with below-average specialisation in the corresponding sector. On the other hand, the closure of existing plants tends to increase it. Unfortunately, the data set used by Dumais *et al.* (1997) does not distinguish the real death of plants from changes in the property rights or their relocation.

More details are available from data on French establishments. This data set, which traces all relocations of establishments between employment areas (defined according to commuting patterns), allows for changes in property rights and separates relocations from new plant creations. A complete description of the data is given by Delisle and Lainé (1998). These data have been studied by Duranton and Puga (1999) for the period 1993–96. Over those 4 years, 258 305 new plants were created in continental France (representing 27.8 per cent of the average stock over this period). At the same time, 23 281 establishments relocated to another employment area (2.5 per cent of the average stock over this period).[7] Of all new plants, 58.6 per cent were created in an employment area with above-median diversity (as measured by the inverse of a Herfindahl index of local employment shares of different sectors). Those numbers confirm that most creations take place in diversified cities. Regarding plant relocations, 93.9 per cent of all relocations across employment areas departed from an employment area with above-median diversity. Two-thirds of those went to locations with above-median specialisation (as measured by the share of the corresponding sector in local employment). This implies that 61.6 per cent of all relocations across employment areas both departed from

an employment area with above-median diversity and arrived in an employment area with above-median specialisation, while less than 3 per cent of all relocations across employment areas were from an employment area with below-median diversity to an employment area with below-median specialisation. Note that this spatial bias in the movement of firms is necessary to make the bias in the creation of new plants towards larger and more diversified cities consistent with the stability in city size and specialisation described under Stylised fact 3. If new firms and new sectors were born in larger and more diversified cities without any posterior differences, those cities would come to dominate the urban landscape, whereas specialised cities should disappear, and we have seen that this is not the case.

Not only is the creation of new plants biased towards larger and more diverse cities, but so is the location of innovative activities that lead to new products. Jacobs (1960) was the first to argue forcefully that localised diversity may favour innovation. Nonetheless, it is only recently that hard evidence on this issue has been gathered. Feldman and Audrescht (1999) use a data set of 3969 US product innovations in 1982 for which the address of the innovating establishment can be identified. They first find that a staggering 96 per cent of the innovations were made in those metropolitan areas, which accounted for only 30 per cent of the US population. Cities are thus the places where innovations occur. But it is important to establish what is the main force driving the innovative process: pure city size, urban diversity or urban specialisation. To investigate this issue, Feldman and Audrescht (1999) relate each four-digit industry to an underlying scientific base (using previous scientific surveys), and group together industries sharing a common scientific base.[8] Using this information, they first show that industries sharing a common scientific base tend to cluster together. Then for each city-industry, they regress the innovative output on an index of relative specialisation in the sector, an index of diversity across industries with a common science

base, city size and other controls. They find that own-industry specialisation has a negative effect on innovative output, whereas diversity across industries with a common science base and city size both have a positive and highly significant effect on innovative output. Similar results are obtained when the same analysis is carried on at the level of the firm. Using data for nine major Japanese electronic firms, Fujita and Ishii (1998) show that their R&D activities are located solely in the major metropolitan areas of Tokyo, Kyoto and Boston. When looking at production, they distinguish between trial plants and mass-production plants, and show that trial plants are also overwhelmingly located in major metropolitan areas, whereas mass-production plants are almost always located in smaller cities or rural areas.

Having outlined the main stylised facts about diversity and specialisation in cities, we now turn to theories looking at these issues.

3. Static Theoretical Approaches with No Trading Costs for Final Goods

Most theories of urban systems dealing with issues of size, diversity and specialisation, have some common structure, even if this is not always apparent. To make comparisons easier, we shall describe each theoretical approach starting from a statement about four basic assumptions. First of all, it is useful to view cities as an equilibrium outcome, whether static or dynamic, which balances agglomeration forces (that create benefits for people and firms locating close to each other) and dispersion forces (that create costs from such closeness). Hence, every consistent theory needs to make an explicit assumption about agglomeration forces (Assumption 1) and an explicit assumption about dispersion forces (Assumption 2). Agglomeration forces give a rationale for the existence of cities, while dispersion forces place limits on city size. With respect to agglomeration and following Hoover (1936), we distinguish between localisation economies (agglomeration forces within individual sectors) and urbani-

sation economies (agglomeration forces across different sectors). We think of both agglomeration (also called centripetal) forces and dispersion (centrifugal) forces as arising from economic interactions. Non-economic interactions, even if important in explaining early cities, are probably not strong enough to sustain today's larger cities. The limits placed by congestion costs on city size imply that not everyone concentrates in a single location. This calls for two additional assumptions. One of these spells out the relationships between cities in an urban system, which requires making assumptions about the mobility of goods and households (Assumption 3). Finally, the structure of the urban landscape needs to be made explicit. Either the number and locations of cities are taken as exogenous, or some mechanism for city formation needs to be proposed (Assumption 4).

The starting-point of the first thread of the literature is Henderson (1974). Building on the Alonso–Muth framework (Alonso, 1964; Muth, 1969), which models the internal structure of cities, he extends it to a system of cities in general equilibrium with a finite number of sectors. The following is a simplified description of his model. He makes four main assumptions.

Assumption 1a (Marshallian agglomeration). Existence of own-sector externalities.

In Henderson (1974), localisation economies take the form of pure Marshallian externalities, whereby the productivity of labour in any given sector in any given city is assumed to increase with total employment in that sector in that city. The strength of local externalities is also assumed to vary, so that these are stronger in some sectors and weaker in others.

Assumption 2a (dispersion through crowding). Cities are subject to congestion costs, which increase with their population.

This can be easily justified with reasonable assumptions about commuting costs and land rents. Consider the simple case of a city with

a single central business district. An extra worker added at the edge of a city will face higher commuting costs than the worker previously furthest from the city centre. This will make closeness to the city centre more valuable, and lead to higher land rents. The combination of commuting costs and land rents is a congestion effect that increases with population.

Assumption 3a (perfect product and labour mobility). Inter-city trade is free for all final goods and labour is perfectly mobile across cities.

The first part of this assumption, the possibility of transporting between cities at no cost, is most important. Relaxing it, as we shall see later in section 4, can qualitatively change the results. By contrast, the extreme assumption of perfect mobility of labour across cities will not play a very big role in this context. In a static framework, some form of labour mobility is necessary for the final landscape not to be fully determined by the initial conditions. At the same time, the degree of this mobility is not crucial.

Assumption 4a (large agents). Large agents can create new cities.

In Henderson (1974), large agents take the form of urban developers. These developers are competitive in the sense that they maximise their profits, taking as given the product and labour-market implications of the behaviour of other developers.

Note that under Assumptions 1a, 2a and 3a, no developer is willing to open more than one sector in a given city. The reason is simple: by segregating sectors across cities, congestion is reduced without affecting productivity, since positive externalities take place only within sectors. Alternatively, by keeping city size constant and specialising the activities of a city, a developer keeps congestion at the same level and increases the productivity in this city. Thus productive efficiency calls for complete urban specialisation, which is implemented by profit-maximising developers. Free entry of developers drives these profits to zero, which

ensures that cities are of optimal size. This equilibrium/optimal city size is the result of a tension between localisation economies, which act as an agglomeration force, and urban crowding, which tends to disperse the population. Equilibrium/optimal city size is increasing with the degree of localisation economies and decreasing with the extent of crowding. Since cities specialised in different sectors experience economies of localisation with a different intensity, they are of different sizes, but cities with the same specialisation are of the same size.

Result 1. Under Assumptions 1a, 2a and 3a, the urban system is only composed of fully specialised cities, provided that the initial number of cities is large enough. Furthermore, imposing Assumption 4a results in an efficient urban system with cities of optimal size.

Here we have presented Assumptions 1a–4a in their most extreme forms for expositional clarity, so that the lack of realism need not be of concern. A more relevant shortcoming is the lack of microeconomic foundations for localisation economies under Assumption 1a. These foundations have a theoretical importance in understanding what localisation economies precisely are, and how these may vary as other features of the urban system change. Such foundations are also important empirically when one tries to measure those externalities. Finally, the precise understanding of local externalities is also crucial from a policy perspective, particularly since the latter often call for corrective policies; one can hardly evaluate the need for, and design of, such corrective policies without an adequate understanding of the nature of the externalities.

Regarding its predictive ability, this model is consistent with the existence of specialised cities of similar size having the same specialisation (Stylised fact 2). However, this model also implies that all cities are fully specialised cities and rules out diversified cities, which contradicts Stylised fact 1. In this framework, provided that cities do not overlap, their relative locations do not matter, and

this can be related to some of the features described under Stylised fact 4. Finally, the model predicts the emergence of a market for cities with active land developers on the supply side. This is a useful construct and one that is very relevant for understanding the formation of many edge cities in the US (Garreau, 1991). It also suggests that restrictions on the ability of land developers to operate may be a contributing factor to the large (often to the extent of being considered inefficient) size of the typical metropolis in less-developed economies. At the same time, it may be less applicable to other geographical areas such as Europe, where land developers play less of a role. Nonetheless, it can be shown that under certain conditions the more plausible assumption of active local governments leads to similar results.[9] Furthermore, Assumption 4a just ensures the optimal city-size result. Even without any large agent, specialisation will occur provided the initial number of cities is large enough.[10]

Overall, this model remains a true landmark in the understanding of urban systems, allowing us to capture in a simple framework a general equilibrium of a system of cities, where the cities arise endogenously from a tension between economies of agglomeration and urban crowding, and interact with each other through trade. A large fraction of the subsequent literature takes this model as its starting-point. In the rest of this section, we shall examine models that try to amend Assumption 1a. Theories dealing with departures from Assumptions 3a and 4a are presented in section 4, whereas more complicated dynamic structures are introduced in section 5.

A number of authors have provided adequate microeconomic foundations for Assumption 1a. Abdel-Rahman (1988), Fujita (1988) and Rivera-Batiz (1988) took this step around the same time, by introducing Chamberlinian monopolistic competition in the standard urban framework. The extension of this line of research to a general-equilibrium model of urban systems can be found in Abdel-Rahman and Fujita (1990). Assumption 1a is replaced by

Assumption 1b (Chamberlinian agglomeration). Final goods are produced by using non-tradeable differentiated services, and a larger range of locally available services lowers the costs of production for final goods.

The details of the assumption are as follows. Final goods are produced in a two-stage process. First, labour is used by monopolistically competitive service firms to produce differentiated producer services under increasing returns to scale. These firms are monopolist producers of their own variety and thus choose prices to maximise profits. However, free entry and exit drive their profits to zero. These services are, in turn, used by competitive final-good producers who have a production function where the different varieties enter with a constant elasticity of substitution. The larger a sector in a city, the larger is the range of service producers that it can support, and the higher the final output that can be produced from the same labour supply. Thus the replacement of Assumption 1a by 1b leads to Result 1 again.

Yet the contribution of Abdel-Rahman and Fujita (1990) is important beyond this equivalence result, in that it suggests a different source for the increasing returns: Chamberlinian economies of agglomeration instead of the pure Marshallian economies in Henderson (1974). Even though the aggregate implications of Marshallian and Chamberlinian economies of agglomeration are the same, Abdel-Rahman and Fujita (1990) show that the policy implications can be very different. Under Assumption 1a (Marshallian agglomeration), the production of each good is subject to constant returns to scale, and firms are perfectly competitive, yet there is a city-wide externality. By contrast, under Assumption 1b (Chamberlinian agglomeration) the production of each intermediate service is subject to increasing returns to scale at the level of each imperfectly competitive firm, and economies of agglomeration stem from price interactions between firms. In their particular model, Abdel-Rahman and Fujita (1990) show that optimal city size can be reached

without much intervention under Assumption 1a. Under Assumption 1b, however, a production subsidy is required to reach optimal city size.

Helsley and Strange (1990) propose an alternative microeconomic foundation for urban externalities, using a labour-market argument. They rely on the general idea that cities may offer specialised workers more employment opportunities and that, conversely, a larger pool of workers makes it easier for employers to find specialised labour. Consequently, the matching between employers and specialised employees is easier in larger cities. Furthermore, cities can also be thought of as fostering the specialisation of labour in the first place. Duranton (1998) shows how the specialisation of labour implies economic agglomeration, which helps to sustain a high degree of labour specialisation (Smithian agglomeration). The main assumption is one of increasing marginal productivity when the time spent on a given range of tasks increases. When final goods are produced using localised specialised services, this again leads to a reduced form similar to those obtained by Henderson (1974) and Abdel-Rahman and Fujita (1990) at the city level. However, Smithian agglomeration economies can be thought to be different in nature from both Marshallian and Chamberlinian agglomeration economies. Smithian agglomeration economies stem from interactions at the level of the labour market and are likely to result, yet again, in different prescriptions in terms of economic policy.[11]

Providing microeconomic foundations for the production function, however useful, still does not allow us to generate both diversified and specialised cities—complete urban specialisation is still the only possible outcome in this thread of literature. Abdel-Rahman (1990a, 1990b), Abdel-Rahman and Fujita (1993) and Abdel-Rahman (1994) provide the four main static models which introduce the possibility that some cities specialise and others diversify. Abdel-Rahman (1990a) amends Assumption 1b slightly by including non-tradeable final goods.

Assumption 1b′ (Chamberlinian agglomeration with non-tradeable sector). Final goods are produced by using non-tradeable differentiated services, and a larger range of locally available services lowers the costs of production for final goods. There is a non-tradeable and essential final-good sector that can engage in production using the same differentiated services as any of the tradeable sectors.

The main result of Abdel-Rahman (1990a) is:

Result 2. Under Assumptions 1b′, 2a, 3a and 4a, all cities manufacture the non-tradeable final good and one of the tradeable final goods. In equilibrium, larger cities are more specialised in the production of one of the tradeable goods.

The first part of the result is easy to understand. Since the non-tradeable good is essential in the utility function, all cities will produce it in equilibrium. The existence of this non-tradeable sector will tend to limit urban specialisation. This is a desirable property of the model, since we saw earlier that most cities had a large part of their labour force employed in sectors such as retail or health services. The other important aspect of the model is that tradeable and non-tradeable final goods can make use of the same intermediate goods. This aims to capture the idea of diversification arising from cost-saving due to shared inputs. The second part of the result is less intuitive. To understand it, consider the following thought experiment. Starting with a city of small size, imagine an increase in its population. Then, due to the economies of agglomeration, production will become more efficient. Thanks to this increase in efficiency of the city, the population will be richer and thus will demand more imported goods. The only way that this can be paid for is by specialising more in the production of the tradeable good. This prediction can be meaningful when distinguishing very small cities, which are often dominated by the production of non-

tradeable goods to a greater extent than medium-sized and large cities. But if our focus is on medium-sized and large cities, as it is here, this result contradicts Stylised fact 2.

An alternative way of generating diversified cities is to introduce cross-sector externalities. This leads to a tension between localisation and urbanisation economies. Abdel-Rahman (1990b) develops this idea in a partial-equilibrium framework. His main assumption is the following.

Assumption 1a′ (Marshallian agglomeration plus cross-sector externalities). Existence of own-sector externalities and of cross-sector externalities.

The simplest way to represent this assumption is to consider only two sectors, with positive own-sector externalities in each sector, and also with a positive cross-sector externality from sector 2 to sector 1. In this partial-equilibrium analysis, the number of cities is fixed (and their location is irrelevant given the absence of trading costs).

Assumption 4b (set landscape). The number and location of cities is fixed.

If the two products are essential in the utility function, we reach the following result.

Result 3. Under Assumptions 1a′, 2a, 3a and 4b, three configurations are possible in equilibrium: only specialised cities (good-1 cities and good-2 cities); only diversified cities; and mixed configurations with both good-2 cities and diversified cities. Under a mixed configuration, diversified cities are smaller than specialised cities.

Regarding the first part of the result, note that if cross-sector externalities are very weak, these can actually be ignored. This drives us back to Assumption 1a, and we obtain Result 1 again, with only specialised cities. In the opposite case, when cross-sector externalities are very strong and own-sector externalities very weak in sector 1, this sector will always be attached to sector 2, and the equilibrium configuration will always imply either only diversified cities or a mixed configuration.

The last part of the result is more subtle. Diversified and specialised cities cannot be of the same size. If they were, diversified cities would have the same urban crowding with weaker localisation economies in sector 2. This would make the net wage of labour employed in sector 2 lower in diversified cities. But, with labour mobility, workers in sector 2 must receive the same net wage in both types of city. Potentially, wage equalisation across cities could be ensured by having sector 2 larger in diversified cities than in specialised cities. But this situation is unstable as individual workers gain by relocating from diversified to specialised cities, where there is less crowding. This would reinforce the economies of agglomeration in specialised cities. Thus diversified cities cannot remain larger than specialised cities. It can be shown formally that the only possible stable equilibrium is to have specialised cities being larger than diversified cities. This, again, is in contradiction with Stylised fact 2. The other empirical limit to the model is that it precludes the co-existence of diversified cities and specialised cities of both types (i.e. good-1 cities, good-2 cities and diversified cities). This type of co-existence, however, seems relatively common. In France, for instance, Sochaux is specialised in the car industry and Toulouse in the aircraft industry, whereas Paris is diversified with both industries being (non-trivially) represented.

Abdel-Rahman and Fujita (1993) attempt to tackle the same problem differently, by taking a general equilibrium approach. They start from the idea that diversification may stem from economies of scope. In their model, they consider again two sectors, 1 and 2, producing homogeneous goods, and three possible technologies for factory cities.

Assumption 1c (agglomeration through fixed costs). Three technologies are available. The first technology involves a labour overhead and a rising marginal cost to produce good 1. The second technology also involves a labour overhead and a rising marginal cost to produce good 2. Finally, the third technology involves a

labour overhead and a rising marginal cost to produce jointly both goods.

Clearly, the use of the first two technologies leads to specialised cities, whereas the last one results in diversified cities. Regarding dispersion, Abdel-Rahman and Fujita (1993) impose Assumption 2a. Perfect mobility is also assumed for final goods (Assumption 3a). Finally, their general equilibrium framework leads to the reintroduction of Assumption 4a. Their main result is:

> *Result 4.* Under Assumptions 1c, 2a, 3a and 4a, three configurations are possible in equilibrium: only specialised cities; only diversified cities; and mixed configurations with both diversified cities and specialised cities of one type. Under a mixed configuration, diversified cities are larger than specialised cities.

Thus, the three types of configuration that emerge in equilibrium are very similar to those in Abdel-Rahman (1990b). However, unlike this latter paper, Abdel-Rahman and Fujita (1993) find that when specialised cities exist, they are smaller in equilibrium than diversified cities. This is easy to understand. If the labour overhead associated with the mixed technology 3 is relatively small, only diversified cities will exist. For a mixed configuration (and thus specialised cities) to be sustainable, a relatively large labour overhead is needed in the diversified city. This will result in diversified cities being of a larger size. This is an attractive result in that it is consistent with Stylised fact 2. What is not so appealing is that, together with the other results from this model, it is driven by large city-specific fixed costs. As a consequence, cities are taken to be factory towns, which are not very common nowadays.

Abdel-Rahman (1994) pursues this line of investigation further, and tries to provide some microeconomic foundations for the previous model. He does so by introducing a production structure with two stages. Service producers manufacture differentiated intermediate goods, which are later aggregated into final goods as under Assumption 1b. The twist here is that service producers have the choice between different technologies as with Assumption 1c (so that large city-specific fixed costs are replaced by the more plausible assumption of small firm-specific fixed costs). Intermediate goods producers can thus choose whether to produce intermediates for any one sector or for both. As before, the same three types of configuration are possible. However, this model brings back the counterfactual result that in the mixed configuration specialised cities are larger than diversified cities. This occurs for the same reason as in Abdel-Rahman (1990b).

All the papers described in this section propose very intuitive explanations as to why there might be a tension between localisation and urbanisation economies, leading cities to diversify or specialise depending on the circumstances. The models used to carry through the analysis are clean and tractable. Nonetheless, this type of approach has some important limits since it seems difficult to take the consistency with stylised facts much beyond the co-existence of diversified and specialised cities of Stylised fact 1.

4. Approaches with Trading Costs for Final Goods

The second strand of literature can be traced back to Christaller (1933) and central-place theory. Christaller starts from the observation that goods are bought with different frequencies, and argues that this may justify the existence of an urban hierarchy. For instance, imagine that good A is bought on a daily basis, whereas good B is bought on a monthly basis. He argues that this should give rise to a hierarchical system of cities, with some cities selling only good A (central places of level 1), and other cities selling both good A and good B (central places of level 2). In a two-dimensional space (and using a specific metric), this gives rise to the famous contiguous hexagonal market areas for central places of level 2. The market areas for the higher-frequency goods are successively built on the market areas of the

lower-frequency goods. Christaller justified this structure on the basis that it would minimise transport costs. The strength of this approach is that the postulated configuration matches real urban landscapes in certain important respects and can be related to Stylised facts 1, 2 and 4 (see Berry and Parr, 1988, for a more detailed review). However, there is no explanation as to why such a configuration would emerge as an equilibrium outcome of decentralised actions by economic agents. The urban system postulated by central-place theory is such that transport costs are minimised, but no one explicitly performs this minimisation, nor does it result from the microeconomic behaviour of firms and consumers. Eaton and Lipsey (1982) were the first to propose an equilibrium model of central places in which they try to deduce the propositions of central-place theory from rigorously stated assumptions. Given the enormous difficulty of the problem, they were only able to present a very stylised model in partial equilibrium. The main assumptions are the following.

Assumption 1c′ (agglomeration through fixed costs with cities as shops). Profit-maximising retailers, facing a fixed cost for each new shop, sell two homogeneous goods, A and B. Prices are exogenously determined.

Compared with Assumptions 1a and 1b, the primary source of agglomeration is the fixed cost faced by retailers, which prevents them from spreading continuously across space. While treating prices as parametric is not very satisfactory, having price competition makes the model of Eaton and Lipsey (1982) intractable.

Assumption 2b (direct dispersion). Households are uniformly distributed over the real line. They consume good A and good B at constant unitary rates.

Note that this assumption is much more extreme than Assumption 2a, since here the dispersion of demand is assumed directly (and in such a way that cities are 'empty' of people) and not derived within the theory.

The focus of the model is solely on shopping issues.

Assumption 3b (shopping costs). Households are allowed to shop at regular intervals. When they shop, they can buy at most one bundle of fixed size of each good. Households shop so as to minimise shopping costs.

Here, the free-trade Assumption 3a is replaced by an assumption of strictly positive trading costs. These trading costs take the specific form of shopping costs, capturing the indivisibilities of shopping (goods are bought in bundles at regular intervals).

Overall Assumptions 1c, 2b, 3b and 4a can be seen as the rigorous statement of the implicit assumptions in the informal story given at the beginning of this section. Beyond this formal statement of the assumptions of traditional central-place theory, Eaton and Lipsey (1982) also make explicit the decision to locate cities. Households choose their shopping strategy so as to minimise shopping costs, while retailers choose their location so as to maximise their profits, taking the action of other retailers as given (using the concept of a Nash equilibrium). From this set of assumptions they derive the following result.

Result 5. Under Assumptions 1c′, 2b, 3b and 4a, at least one central place of level 2 exists in equilibrium. If central places of levels 1 and 2 co-exist, then all central places of level 1 offer the same good.

The first part of the result requires the following thought experiment. Imagine that a central place of level 1 which sells good A is located to the left of a central place of level 1, which sells good B. Then the retailer selling good A would gain by shifting rightwards and relocating closer to the retailer selling good B. By so doing, the retailer would benefit from the 'lump' of demand which went to the shop selling good B (these benefits are often called 'economies of multi-purpose shopping'). Similarly, the retailer selling good B will benefit from a leftwards move. The benefits from the two retailers

locating closer to each other will only be exhausted once a central place of level 2, selling both goods, is created. This central place of level 2 allows each retailer to take advantage of shoppers going to the other retailer. In short, each store creates a positive demand externality for stores selling the other good.

The previous reasoning implies that a central place of level 1 selling good A cannot locate next to a central place of level 1 selling good B. Thus, the equilibrium must have either only central places of level 2, or central places of level 2 alternating with central places of level 1 all selling the same good. This result confirms some of Christaller's intuitions and provides justifications for aspects of Stylised facts 1, 2 and 4. In particular, the physical location of cities can now be considered in the model, albeit in a very stylised way. However, it also shows that the typical central-place pattern need not emerge in equilibrium, and that the regular geography proposed by Christaller does not generally arise. In terms of the economic underpinnings, the insights of a central-place theory driven by shopping costs can help to illuminate many issues surrounding small cities, but they are probably of much less relevance for larger cities producing goods sold everywhere in the economy. This suggests replacing the assumption of shopping costs by one of shipping costs.

Assumption 3c (shipping costs). Shipping goods to consumers involves costs that increase with distance and differ across goods.

This then leads simply to:

Result 6. Under Assumptions 1c′, 2b, 3c and 4a, there is no tendency for producers to cluster.

The tendency of producers to cluster disappears because, without shopping costs, there are no economies of multipurpose shopping. The derivation of the equilibrium market area of firms depends on the pricing mechanism. But it can be shown that in each sector, competition will lead to an equilibrium market area where producers make zero profits. The equilibrium market area differs from sector to sector, but there is no reason for the ratio of the size of market areas in different sectors to be exactly an integer, so that central places can only arise by chance (Anderson *et al.*, 1992). The difficulty of dealing with physical distance using well-specified models, along with the negative results derived by Eaton and Lipsey (1982), drastically limited the amount of research using this type of approach. The replication of the regular geometry of central-place theory started to be viewed as a nearly impossible task. It is only in the very recent past that new models have been developed, in which physical space is considered explicitly.

Coming from a radically different perspective from that of Eaton and Lipsey (1982), Abdel-Rahman (1996) develops a model similar in spirit to Henderson (1974). The initial framework is the same, but he introduces trading costs and fixes the initial location of cities (Assumption 4b). In his setting, there are two sectors in which final goods are produced by using labour and differentiated intermediate goods (Assumption 1b). Cities are monocentric with congestion due to crowding (Assumption 2a). Finally, the costs of shipping goods between any two cities are the same.

Assumption 3c′ (shipping costs and perfect labour mobility). Shipping goods to consumers involves costs that increase with distance. Labour is perfectly mobile.

This assumption is very close to Assumption 3c. It can also be thought of as Assumption 3a, but with shipping costs being strictly positive. This allows Abdel-Rahman to study the conditions under which cities specialise or diversify. His main result is:

Result 7. Under Assumptions 1b, 2a, 3c′ and 4b, two configurations are possible in equilibrium: pure specialisation and pure diversification. Pure specialisation arises for low shipping costs, while pure diversification arises for high shipping costs.

It is useful to compare this result with Result 1. Without shipping costs, cities specialise as in Henderson (1974). Again, diversification is costly since it reduces the economies of localisation while holding urban crowding constant. Here, however, positive shipping costs also make specialisation costly, since it is expensive to import from other cities. As a result, when shipping costs are high, cities will diversify. This result can be thought of as capturing the evolution of cities in the past. During the Industrial Revolution, lower shipping costs coincided with the rise of more specialised cities (Bairoch, 1988). However, in the more recent past, the decline in shipping costs does not seem to have promoted more specialised cities. On the contrary, large diversified cities seem to have done particularly well (Black and Henderson, 1998; Glaeser *et al.*, 1992). Note also that the model does not predict the co-existence of diversified and specialised cities (in contradiction with Stylised fact 1). Finally, the modelling of physical space remains 'incomplete', since the model considers only the impact of spatial frictions on the composition of cities with their location exogenously given.

The next step towards a better modelling of physical distance in urban systems was taken by Fujita and Krugman (2000) and Fujita *et al.* (1999), who adapted the approach developed by Krugman (1991) for regional systems to a more specifically urban framework. Their approach represents a significant departure from central-place theory in that they do not try to generate a regular urban system. According to Christaller, existing urban systems are irregular because of physical features and other distortions. Remove the distortions and urban systems should tend towards geometric regularity. By contrast, Fujita and Krugman (2000) show that lack of perfect regularity in the spatial distribution of cities may be a fundamental property of urban systems, which would arise even without any unevenness in physical space. The general-equilibrium model built by Fujita and Krugman (2000) relies on the following assumptions.

Assumption 1b″ (Chamberlinian agglomeration from consumer tastes). There is a single manufacturing sector producing differentiated consumer goods. Consumers have a taste for variety.

Note that this assumption is very close to Assumption 1b, except that the differentiated goods are directly consumed instead of being used to produce another good. The important aspect of consumer preferences regarding manufactured goods is the taste for variety.

Assumption 2c (dispersion through fixed factors). Agricultural land is represented by a line. The agricultural good is essential. Agricultural income is spent where it is generated.

This assumption introduces a second type of good: a homogeneous agricultural product, to which consumers devote a fraction of their income. Unlike the differentiated goods that give rise to localised increasing returns, the agricultural good is produced under constant returns to scale. Another crucial difference between agricultural and manufactured goods is that the production of the former requires not only labour but also an immobile factor (land), and returns to this factor follow the spatial distribution of land.

Regarding mobility, Fujita and Krugman (2000) assume again that shipping final goods is costly and that labour is mobile (Assumption 3c′). City formation is now the aggregate outcome of individual decisions by firms and workers.

Assumption 4c (self-organisation). Cities can form endogenously from the optimising behaviour of individual firms and workers.

An equilibrium is a situation in which the following obtain: workers maximise their utility and firms their profits, subject (respectively) to the relevant budget and technological constraints; individual workers are unwilling to relocate; it is unprofitable for more firms to begin production; and no existing firm wants to relocate or cease pro-

duction. These assumptions lead to the following result:

> *Result 8.* Under Assumptions 1b″, 2c, 3c′ and 4c, there are many possible equilibrium configurations. In particular, a system of cities with cities of different sizes can arise endogenously.

Consider first how the propensity of firms and workers to agglomerate arises in this context. As is usual with Chamberlinian monopolistic competition, each variety will be produced by a single firm. A larger number of locally produced varieties of manufactures will make a location more attractive to consumers (all goods can be consumed everywhere, but it is less costly to buy them closer to the place of production). Migration will increase market size, making the location more attractive to firms. This creates a mechanism of circular causation, by which more firms attract more workers/consumers and more workers/consumers attract more firms. Note that the intuition is similar to Abdel-Rahman and Fujita (1990) which is discussed above, the main difference being that here the model deals with imperfectly tradeable final goods instead of non-tradeable intermediate goods.

However, it is not always the case that all firms agglomerate at one location. When the economy is large enough, a unique city may not be sustainable in equilibrium. Since manufactured goods are substitutes, some firms may find it profitable to relocate away from the city and serve mainly farmers located in the agricultural fringe. This quite naturally generates a system of cities, with cities of different size but with smaller cities being miniature replicas of the larger cities. To show this result, the analytical tool used by Fujita and Krugman (2000) is the market-potential function. This tool has a long tradition in economic geography. However, they provide a microeconomic basis for it, regarding it as the profit at every possible location for a firm considering relocation. They further refine it by taking into account substitution effects between goods. This market potential has two main components: demand from the city and demand from the

farmers. As a firm moves away from the city, the first component declines while the second increases. The sum of these two components of the market potential does not change monotonically as distance to the city increases. Close to the city, the negative effect of losing urban demand is stronger; but, further away, the positive effect of stronger agricultural demand may dominate. Thus, typically, there will be a large city in the centre of the economy and smaller cities close to the fringes.

When Fujita and Krugman (2000) introduce the possibility of multiple sectors with goods having different shipping costs, they can generate hierarchical systems of cities as in traditional central-place theory. Starting from an economy with a single city, a growth in population will expand the agricultural hinterland and lead sectors with higher shipping costs to spread out. The original city will keep all sectors, but new peripheral cities will form, initially based on those goods with high shipping costs and importing the rest from the original city. Thus, in equilibrium, it is possible to obtain a system of cities with large and diversified cities along with small and specialised cities. This modelling framework is quite attractive although analytical complexity and multiplicity of equilibria emerge as two important drawbacks. To deal with these issues, Fujita *et al.* (1999) propose a refinement of the previous approach. They use an evolutionary framework in which population grows over time and they assume a set of simple and plausible rules for equilibrium selection. Using numerical simulations for a three-sector economy, they show that population growth leads to the formation of a hierarchical system of cities with large and diversified cities and small specialised cities, as in Fujita and Krugman (2000). They also show that typically central places of level 1 will be located between central places of higher levels which offer a broader range of goods.

The work of Fujita *et al.* (1999) represents a landmark in the modelling of urban systems, since this is the first paper able to deal with central-place theory in a general-

equilibrium framework with well-specified microeconomic foundations. Empirically, it can again replicate well Stylised facts 1 and 2. It also provides some spatial predictions related to Stylised fact 4. Nonetheless, this approach has certain drawbacks. First, the model yields very few testable results. Compared to Fujita and Krugman (2000), it can tackle the multiplicity of equilibria reasonably well. However, it still relies on numerical results, and it would appear that almost any observed urban system can be seen to be a possible outcome of their model. The second undesirable property of their model is the constant creation and destruction of new cities. When a new city is formed or when transport costs decline, the whole urban system is often reshuffled. Perhaps this excessive urban turbulence would be mitigated in a more realistic framework, with mobility costs for labour and sunk capital investments. The third problem is that the entire model relies on the differences in market potential between a continuum of locations. This introduces physical space into the model in a meaningful way, but the approach may be too extreme since, as discussed under Stylised fact 4, the evidence on market potential driving urban growth is relatively weak. Finally, this approach relies purely on self-organisation. While Assumption 4a assigns too strong a role to land developers, here large agents (including central or local government) are ignored completely. More work in the middle ground is clearly needed.

5. Dynamic Approaches

So far, we have discussed essentially static theoretical approaches to the study of urban systems. These necessarily ignore some of the most important stylised facts reviewed in section 2—those describing the dynamics of urban systems and of economic agents in such systems. There is a significant tradition of dynamic urban modelling, surveyed by Miyao (1987), but the focus of most of this literature rests solely on issues of urban size and residential investment. It is only very recently that dynamic theoretical models

have been proposed to try to understand better some of the stylised facts about specialisation and diversity in cities.

Brezis and Krugman (1997) attempt to explain why particular sectors may rise and fall in some cities (and, in certain cases, where the city is heavily specialised, draw the city along in this rise and fall). While we saw that the overall picture is one of stability in the production structure of cities (Stylised fact 3), this does not preclude significant changes from time to time. Brezis and Krugman (1997) propose a simple partial-equilibrium model of a region with one initial city, and focus on the dynamics of localisation economies.

Assumption 1d (learning-by-doing and localised knowledge leading to agglomeration). Production technologies are localised. New technologies in a given sector arise randomly. Social learning-by-doing implies that a new technology may initially be less productive than the old one but will eventually become more productive.

Brezis and Krugman (1997) assume an initial state of the world where a dominant technology is available to manufacture a given good. This dominant technology is localised in an established city which specialises in the manufacture of the good. Since knowledge is localised, the good cannot be profitably produced elsewhere using the dominant technology. There is also social (in the sense that no firm or worker is large enough to internalise its effects) learning-by-doing. This increases the productivity of labour with the amount of accumulated production in the city where the dominant technology is used. Brezis and Krugman (1997) also assume that cities are subject to crowding (Assumption 2a). Finally, for the sake of simplicity, they also make Assumption 3a about the perfect mobility of goods and labour. Their main conclusion is Result 9.

Result 9. Under Assumptions 1d, 2a, 3a and 4c, the arrival of a new technology can lead to the gradual emergence of a new

city and the fall of the established city using the old technology.

This urban leapfrogging arises because, when a new technology is introduced, firms in the established city find it preferable to keep using the old technology at a higher level on the learning curve. Since learning-by-doing is collective, no firm in the established city is willing to start using it and pay the cost of a learning-by-doing process, the benefits of which will be freely available to all. However, urban crowding in the established city may make it profitable for some workers to break away and form a new city. There, the lower initial productivity of the new technology may be offset by the absence of urban crowding. In equilibrium, the allocation of workers across the new and the established city is such that wages are equalised across locations. This implies that the new city is initially much smaller than the established city since, with its less-productive technology, it cannot support crowding costs as large as those for the established city. However, the potential for learning-by-doing in the new technology is more important. Over time, the productivity of labour in the new city will rise relative to that in the established city. This will induce workers to migrate to the new city and reinforce learning by doing until the production of the good is entirely concentrated in the new city. Brezis and Krugman (1997) provide an elegant formalisation justifying the episodes of changing urban fortunes described as exceptions to Stylised fact 3. While the assumptions about learning by doing are rather special and extreme, Amiti (1998) shows that a similar leapfrogging result holds with Chamberlinian economies of agglomeration.

Duranton and Puga (1999) take a broader view about the dynamics of firms and products within urban systems. Their framework is related to those of Henderson (1974) and Abdel-Rahman and Fujita (1990), but introduces an explicit dynamic dimension related to the process of firms learning how best to produce new goods.

Assumption 1b‴ (Chamberlinian agglomeration and imperfect initial knowledge). A firm making a new good does not know what the best production process is. It can produce prototypes with processes drawn from existing local sectors. A firm recognises its ideal process after producing a prototype with it, and is then able to begin mass production. A larger number of firms using the same type of process lowers production costs associated with that type of process due to Chamberlinian agglomeration economies. Firms have some probability of closing down in any period.

As in most models reviewed, cities are subject to urban crowding (Assumption 2a). Assumption 3d summarises mobility issues.

Assumption 3d (imperfect firm and labour mobility and perfect product mobility). Firms can change location at a cost. Final goods are freely tradeable. At every period, a small proportion of workers can freely relocate.

When it is costly for firms to relocate, the presence of a learning process that draws from local processes creates dynamic advantages based on diversity. This is because a city with a wider range of local sectors using different processes allows firms producing new products to try more processes in search of their ideal one without incurring relocation costs. At the same time, there are advantages to specialisation stemming from more traditional Chamberlinian agglomeration economies. Urban crowding places a limit on city size and makes the choice between diversity and specialisation a meaningful one. Result 10 summarises the solution to this trade-off.

Result 10. Under Assumptions 1b‴, 2a and 3d, three configurations are possible in equilibrium: specialised cities only; diversified cities only; and a mixed configuration with both specialised and diversified cities. The mixed configuration arises when learning is moderately difficult. In the mixed configuration, firms begin producing prototypes in a diversified

city and relocate to a city of the relevant specialisation to begin mass production when they learn their ideal process. Further imposing Assumption 4a results in an efficient urban system with cities of optimal size. When all three configurations are possible, the mixed configuration dominates the other two in terms of welfare.

In the mixed equilibrium with both diversified and specialised cities, locating in a diversified city can be seen as an investment. It implies higher production costs but also faster learning, which makes possible an earlier transition to more efficient mass production. Production costs are higher in a diversified city because, for a firm using a given production process, the presence of producers using different processes causes urban crowding without creating stronger static localisation economies. However, bearing these higher costs can be worthwhile for new firms because they expect to have to try a variety of processes before finding their ideal one, and a diversified city allows them to do so, without relocating after each trial. Once a firm finds its ideal production process, there are no further benefits from being in a diverse environment. The firm avoids the congestion imposed by the presence of other sectors by relocating to a city where all other firms share its specialisation.

The mixed configuration arises for moderately difficult learning, because if learning is too difficult the investment of initial location in diversified cities does not pay off. On the other hand, if learning is very easy, firms expect not to have to try too many different processes, and prefer to exploit static localisation economies from the beginning. In this dynamic steady state, diversified and specialised cities co-exist. Cities are stable in their size and sectoral composition, but there is a constant turnover of firms. Some existing firms close down during every period, and new firms enter to replace them with new products. This turnover follows a life-cycle pattern. New products are created and developed in diversified cities, but production eventually relocates to specialised cities.[12]

All of this is consistent with Stylised facts 1, 3 and 5. At the same time, the model has less to say about Stylised facts 2 and 4. The optimal size of diversified and specialised cities is the same, and whether specialised or diversified cities are larger in equilibrium depends on the ability of large agents to operate in each type of city. The simplifying assumption of goods being freely traded also ignores the issue of the relative location of cities.

6. Policy Implications and Conclusions

Two main issues arise time and again in the empirical and theoretical work on the structure of systems of cities reviewed in this paper. The first major issue is the efficiency of the urban system as a whole. Existing theories are divided on this point. Models in the spirit of Henderson (1974) tend to suggest that urban systems without central regulation will be efficient in terms of size and number of cities of each type. This is true, provided that some mechanism enabling the creation and development of new cities is present. The two usual mechanisms are autonomous local governments or the market, with land developers. Of course, there are likely to be distortions such as imperfect competition between developers or political-economy issues, making the interests of local communities not perfectly aligned with those of the ruling politicians. Nevertheless, these two mechanisms should bring urban systems reasonably close to efficiency. In contrast, theories where distance matters tend to have a more reserved judgement about the overall efficiency of the urban system. Unfortunately, detailed welfare analysis has not yet been carried out in this type of framework. Despite this division, a few policy prescriptions emerge. For instance, anything that hinders the creation and development of new cities, such as poorly defined property rights, is likely to result in inefficient urban systems. The desirability of other policy actions is less clear. For instance, the need for (and the type of) intervention in the presence of localisation and urbanisation economies depends

delicately on the source of these economies—something that is far from well understood.

The second major issue is about the composition of economic activity in individual cities. From our review, it seems that specialisation has both advantages and disadvantages. The advantages are less urban crowding and stronger localisation economies arising from the proximity of closely related producers. The disadvantages are less innovation and more exposure to risk as the fortunes of specific sectors and technologies rise or fall. Overall, there appears to be a need for both large and diversified cities and smaller and more specialised cities. The main problem with this dichotomy of diversified and specialised cities is that the latter are exposed to a greater risk with the rise and fall of specific sectors and technologies. To limit the risks of urban specialisation, policy may want to curb this. But limiting specialisation may forego important present benefits for the sake of avoiding a possible future downturn. Insurance by the central government may be a good alternative. However, this is costly, and complete insurance may lead to complacency. Another alternative is to give local governments the power to restructure. The disadvantage here is the involvement of local government in excessively ambitious economic-development policies. In particular, not every community can create the next Silicon Valley in its backyard. Moreover, the link between innovation and diversity seems fairly robust, so that highly innovative clusters cannot be bred in previously highly specialised environments. An alternative solution is to encourage labour-force mobility—but in Europe this awakens fears about lost regional identities.

In the end, there will always be a need for large diversified cities like London, New York or Paris, and for more specialised cities like Sheffield or Clermont-Ferrand. The likes of London may always be more dynamic than the likes of Clermont-Ferrand, but this does not imply that one type of city is economically more desirable than the other. For manufacturing and services, unlike for agriculture, 'sowing' and 'reaping' can take place in different locations. Some cities specialise in churning new ideas and new products; this requires a diversified base, as some of the theories reviewed suggest—and which the evidence appears to confirm. However, other cities specialise in more standardised production; this in turn, is better carried out in a more specialised environment. Urban systems may thus have an innate tendency to create this type of imbalance.

Notes

1. For instance, narrow specialisation is often seen as a means to spur innovation. Yet empirical evidence and some of the theoretical approaches reviewed in this paper show that, in general, narrow specialisation hinders innovation, while diversity across a range of related activities promotes it.

2. Considering all private employment would give similar results, but would lead to the inclusion among the top places of cities specialised in mining such as Duluth, and 'leisure cities' such as Las Vegas and Atlantic City.

3. Since the primary focus of the review is on the patterns of economic activities in cities, we will not deal here with issues that have mostly to do with size such as the rank–size distribution and the growth of average city size, even though we feel that many of these issues are ultimately related to the composition of economic activity.

4. In 1950, Detroit was more than 17 times bigger than Phoenix, but was overtaken by the latter shortly after 1990. Note that we use here population estimates for the 'urban places'. Using a broader definition of cities, such as metropolitan area, would nonetheless yield comparable results, indicating an even stronger growth for Phoenix and a less dramatic decline for Detroit and Pittsburgh.

5. The detailed results of Henderson (1998) are of particular interest. He shows that high-tech industries tend to be more concentrated than machinery-related industries and that they also tend to be more mobile.

6. In Black and Henderson (1998), the market potential of a given city is defined as the sum over each other city of its population divided by its distance to the city under consideration. This is discussed further in section 4.

7. This, of course, underestimates the number of goods whose production relocated, since it

ignores the opening of new establishments that coincide with the gradual phasing out of old ones. Furthermore, many multi-establishment firms do relocate the production of some of their products without actually opening new plants or closing down existing ones. Fujita and Ishii (1998) provide evidence of this from the electronics industry.

8. For instance, the agro-business cluster has 21 industries, and the critical knowledge for this cluster comes from chemistry, agricultural science, computer science and biology.

9. Henderson and Becker (2000) offer an extended comparison of land developers and active autonomous local governments. Helsley and Strange (1997, 1998) provide an in-depth discussion of private land developers and the impact of market imperfections upon land development and urban efficiency.

10. It can be shown formally that if the economy is not constrained by the number of cities, then in the absence of land developers, equilibria with incomplete urban specialisation are unstable.

11. Note that our three categories of agglomeration economies (Marshallian, Chamberlinian and Smithian) reproduce the categories employed by Marshall (1890), since he spoke about the circulation of ideas, the availability of intermediate goods and labour-market pooling. However, a detailed discussion about the nature of externalities is beyond the scope of this paper.

12. Note that the idea of diversified cities being nurseries for new firms and new products is different from the traditional 'incubator hypothesis', which highlights differences between the inner city and its periphery. The traditional focus is on firm creations and movements of firms within a given city and not between cities. The literature on this is discussed by Fagg (1980).

References

ABDEL-RAHMAN, H. (1988) Product differentiation, monopolistic competition and city size, *Regional Science and Urban Economics*, 18, pp. 69–86.

ABDEL-RAHMAN, H. (1990a) Sharable inputs, product variety, and city sizes, *Journal of Regional Science*, 30, pp. 359–374.

ABDEL-RAHMAN, H. (1990b) Agglomeration economies, types, and sizes of cities, *Journal of Urban Economics*, 27, pp. 25–45.

ABDEL-RAHMAN, H. (1994) Economies of scope in intermediate goods and a system of cities, *Regional Science and Urban Economics*, 4, pp. 497–524.

ABDEL-RAHMAN, H. (1996) When do cities specialize in production?, *Regional Science and Urban Economics*, 26, pp. 1–22.

ABDEL-RAHMAN, H. and FUJITA, M. (1990) Product variety, Marshallian externalities, and city sizes, *Journal of Regional Science*, 30, pp. 165–183.

ABDEL-RAHMAN, H. and FUJITA, M. (1993) Specialization and diversification in a system of cities, *Journal of Urban Economics*, 33, pp. 159–184.

ALONSO, W. (1964) *Location and Land-use: Towards a General Theory of Land Rent.* Cambridge, MA: Harvard University Press.

AMITI, M. (1998) *Regional specialisation and technological leapfrogging*. Department of Economics, La Trobe University (mimeograph).

ANDERSON, S., DE PALMA, A. and THISSE, J.-F. (1992) *Discrete Choice Theory of Product Differentiation*. Cambridge, MA: MIT Press.

BAIROCH, P. (1988) *Cities and Economic Development: From the Dawn of History to the Present.* Chicago, IL: University of Chicago Press.

BARTIK, T. (1996) Strategies for economic development, in: R. ARONSON and E. SCHWARTZ (Eds) *Management Policies in Local Government Finance*, pp. 287–311. Washington, DC: International City/County Management Association for the ICMA University.

BERRY, B. J. L. and PARR, J. B. with EPSTEIN, G., GHOSH, A. and SMITH, R. (1988) *Market Centers and Retail Location: Theory and Applications*. Englewood Cliffs, NJ: Prentice Hall.

BLACK, D. and HENDERSON, J. V. (1998) *Urban evolution in the US*. Department of Economics, London School of Economics (mimeograph).

BREZIS, E. and KRUGMAN, P. (1997) Technology and the life-cycle of cities, *Journal of Economic Growth*, 2, pp. 369–383.

CHRISTALLER, W. (1933/1966) *Die Zentralen Orte in Süddeutschland [Central Places in Southern Germany]*, trans. by C. W. BASKIN. Englewood Cliffs, NJ: Prentice-Hall.

DAVIS, S. and HALTIWANGER, J. (1992) Gross job creation, gross job destruction, and employment reallocation, *Quarterly Journal of Economics*, 107, pp. 819–863.

DELISLE, J.-P. and LAINÉ, F. (1998) Les transferts d'établissements contribuent au desserrement urbain, *Economie et Statistique*, 311, pp. 91–106.

DUMAIS, G., ELLISON, G. and GLAESER, E. (1997) *Geographic concentration as a dynamic process.* Working Paper 6270, NBER.

DUNNE, T., ROBERTS, M. and SAMUELSON, L. (1989a) The growth and failure of U.S. manufacturing plants, *Quarterly Journal of Economics*, 104, pp. 671–698.

DUNNE, T., ROBERTS, M. and SAMUELSON, L. (1989b) Plant turnover and gross employment flows in the U.S. manufacturing sector, *Journal of Labor Economics*, 7, pp. 48–71.

DURANTON, G. (1998) Labor specialization, transport costs, and city size, *Journal of Regional Science*, 38, pp. 553–573.

DURANTON, G. and PUGA, D. (1999) *Nursery cities: urban diversity, process innovation and the life-cycle of products*. Department of Economics, University of Toronto (mimeograph).

EATON, C. and LIPSEY, R. (1982) An economic theory of central places, *Economic Journal*, 92, pp. 56–72.

EATON, J. and ECKSTEIN, Z. (1997) Cities and growth: theory and evidence from France and Japan, *Regional Science and Urban Economics*, 27, pp. 443–474.

ELLISON, G. and GLAESER, E. (1997) Geographic concentration in US manufacturing industries: a dartboard approach, *Journal of Political Economy*, 105, pp. 889–927.

FAGG, J. (1980) A re-examination of the incubator hypothesis: a case study of Greater Leicester, *Urban Studies*, 17, pp. 35–44.

FELDMAN, M. and AUDRETSCH, D. (1999) Innovation in cities: science-based diversity, specialization and localized competition, *European Economic Review*, 43, pp. 409–429.

FUJITA, M. (1988), A monopolistic competition model of spatial agglomeration, *Regional Science and Urban Economics*, 8, pp. 87–124.

FUJITA, M. and ISHII, R. (1998) Global location behavior and organizational dynamics of Japanese electronics firms and their impact on regional economies, in: A. CHANDLER, P. HAGSTRÖM, and O. SÖLVELL (Eds) *The Dynamic Firm: The Role of Technology, Strategy, Organization and Regions*, pp. 343–383. Oxford: Oxford University Press.

FUJITA, M. and KRUGMAN, P. (2000), A monopolistic competition model of urban systems and trade, in: J. M. HURIOT and J. THISSE (Eds) *Economics of Cities*. Cambridge: Cambridge University Press (forthcoming).

FUJITA, M., KRUGMAN, P. and MORI, T. (1999) On the evolution of hierarchical urban systems, *European Economic Review*, 43, pp. 209–251.

GARREAU, J. (1991) *Edge city: Life on the New Frontier.* New York: Doubleday.

GLAESER, E., KALLAL, H., SCHEINKMAN, J. and SCHLEIFER, A. (1992) Growth in cities, *Journal of Political Economy*, 100, pp. 1126–1152.

HARRIS-DOBKINS, L. and IOANNIDES, Y. (1998) *Spatial interactions among US cities*. Department of Economics, Tufts University (mimeograph).

HELSLEY, R. and STRANGE, W. (1990) Matching and agglomeration economies in a system of cities, *Regional Science and Urban Economics*, 20, pp. 189–212.

HELSLEY, R. and STRANGE, W. (1997) Limited developers, *Canadian Journal of Economics*, 30, pp. 329–348.

HELSLEY, R. and STRANGE, W. (1998) Private government, *Journal of Public Economics*, 69, pp. 281–304.

HENDERSON, J. V. (1974) The sizes and types of cities, *American Economic Review*, 64, pp. 640–656.

HENDERSON, J. V. (1988) *Urban Development: Theory, Fact and Illusion*. Oxford: Oxford University Press.

HENDERSON, J. V. (1997a) Externalities and industrial development, *Journal of Urban Economics*, 42, pp. 449–470.

HENDERSON, J. V. (1997b) Medium size cities, *Regional Science and Urban Economics*, 27, pp. 583–612.

HENDERSON, J. V. (1999) *Marshall's economies.* Working Paper 7358, NBER.

HENDERSON, J. V. and BECKER, R. (2000) Political economy of city sizes and formation, in: J. M. HURIOT and J. THISSE (Eds) *Economics of Cities*. Cambridge: Cambridge University Press (forthcoming).

HENDERSON, V., KUNCORO, A. and TURNER, M. (1995) Industrial development in cities, *Journal of Political Economy*, 103, pp. 1067–1090.

HOOVER, E. (1936) *Location Theory and the Shoe and Leather Industries*. Cambridge, MA: Harvard University Press.

IOANNIDES, Y. and OVERMAN, H. G. (1998) *Spatial evolution of US cities*. Department of Geography, London School of Economics (mimeograph).

JACOBS, J. (1960) *The Economy of Cities*. New York: Random House.

KIM, S. (1995) Expansion of markets and the geographic distribution of economic activities: the trends in US regional manufacturing structure 1860–1987, *Quarterly Journal of Economics*, 110, p. 881–908.

KRUGMAN, P. (1991) Increasing returns and economic geography, *Journal of Political Economy*, 99, pp. 483–499.

MARSHALL, A. (1890) *The Principles of Economics.* New York: Macmillan.

MIYAO, T. (1987) Dynamic urban models, in: E. MILLS (Ed.) *Handbook of Regional and Urban Economics*, *Vol. II*, pp. 877–925. Amsterdam: Elsevier Science Publishers (North-Holland).

MUTH, R. (1969) *Cities and Housing.* Chicago, IL: University of Chicago Press.

RIVERA-BATIZ, F. (1988) Increasing return, monopolistic competition and agglomeration economies in consumption and production, *Regional Science and Urban Economics*, 8, pp. 125–153.