CENTRO STUDI LUCA D'AGLIANO

WWW.DAGLIANO.UNIMI.IT

CENTRO STUDI LUCA D'AGLIANO
DEVELOPMENT STUDIES WORKING PAPERS

N. 418

January 2017

The changing distribution of firms and workers across cities

*Diego Puga**

* CEMFI and CEPR

ISSN 2282-5452

The changing distribution of firms and workers across cities

Diego Puga[*§]
CEMFI and *CEPR*

September 2015

Abstract: This paper explores the changing distribution of firms and workers across cities. Cities have become increasingly specialized by function rather than sector, with business services, management and innovation concentrating in large urban areas and manufacturing dispersing across smaller cities instead. Big cities attract more educated workers, but the main factor distinguishing big cities is not so much who goes there but what happens to them as a result: locating in a big city allows firms to become more productive and workers to increase their earnings. And yet, as cities have become more different, they have also become more interdependent with important life-cycle patterns emerging. Many new firms and products start up in big diverse cities but eventually relocate, making up space for the next generation of new firms and products. Workers are often willing to pay the higher living cost of bigger cities not just to get higher wages today, but in the hope of learning and acquiring skills that, depending on their own characteristics and luck, they may exploit there or somewhere else.

Key words: cities, heterogeneity

jel classification: r14

*This paper has been prepared within the Value Added in Motion (vam) project funded by the Enel Foundation.
§cemfi, Casado del Alisal 5, 28014 Madrid, Spain (e-mail: diego.puga@cemfi; website: http://diegopuga.org).

1. Introduction

Over the course of history, the locations and sizes of cities have been greatly influenced, first, by land fertility, then, by access to natural and manmade routes of transportation, and subsequently, by proximity to energy sources and primary materials. All of these considerations continue to matter, but much less than before. Natural advantages have gained an important role as amenities but have lost relevance for the location of production. Firms and workers care more and more about the number and characteristics of other firms and workers located near them. This self-reinforcing nature of location (more firms and workers prefer to be wherever more firms and workers are) favours persistence but also opens the door to instability and sudden change.

Over the last few decades, we have seen important changes that have made cities increasingly different from each other. Given that many of these changes are systematically related to city size, bigger urban areas now differ from smaller cities along several important dimensions. In particular, the activities that take place in cities of different sizes, and the characteristics of firms and workers that carry out those activities, are systematically different. Some of these differences are the result of firms and workers with certain characteristics being more likely to locate in big cities than others. Other differences arise because choosing to locate in big cities has consequences that make firms and workers evolve differently. And yet, as big and small cities have become more different, they have also become more interdependent. As a result, important life-cycle patterns have emerged for firms as well as for workers.

This paper reviews recent research exploring the changing distribution of firms and workers across cities and the main implications of these transformations. The review focuses on mature cities in developed countries and the movement of firms and workers between them. We do not cover rural-urban migrations (largely a choice of scope for this review) or the role of cities in developing countries (in this case more a reflection of the current focus of research in urban economics). Fortunately, there are signs that this bias towards developed economies is starting to become less extreme through a combination of improved micro data availability for developing economies and ingenious research strategies for careful identification.

We begin in section 2 by reviewing the changing nature of urban specialization, whereby cities have become increasingly specialized by function rather than sector, with business services, management and innovation concentrating in large urban areas and manufacturing dispersing across smaller cities instead. Section 3 turns to differences across firms in terms of their productivity, and the evidence showing the productivity advantages of bigger cities arise because locating close to other firms raises individual firm productivity. Bigger cities also tend to host more educated and particularly productive workers, and section 4 reviews patterns of worker location and their consequences. While increasingly different, cities are also becoming more interdependent, and section 5 discusses this and the life-cycle patterns that are emerging as a result for both firms and workers. Finally, section 6 provides a summary and concludes.

2. The changing nature of urban specialization

Traditionally, the difficulties associated with managing businesses from far away made most firms keep their headquarters and management offices close to their factories. This has gradually changed. Key technological developments in transport and communication technologies combined with new management practices have greatly reduced the additional costs associated with coordinating and monitoring facilities at multiple locations (Chandler, 1977, Kim, 1999). This in turn has profoundly changed the nature of urban specialization, from a situation where cities differed mainly in terms of their sectors of activity to one where cities differ substantially by function (Duranton and Puga, 2005).

The cost of transporting goods, people, and ideas has declined dramatically over the last century. The price of air freight fell from $3.87 to under $0.30 per ton-kilometre between 1955 and 2003, expressed in 2000 US dollars (Hummels, 2007). Revenue per passenger mile for domestic passenger air travel within the United States declined by 70% in real terms between 1960 and 2012 (Bureau of Transport Statistics, 2012). The cost of a three-minute telephone call from New York to London fell from $350 in 1931 to just a few cents today, again expressed in 2000 US dollars (World Bank, 2009). According to Yates (1989, 1991), managerial and accounting innovations have played an even greater role than the reduction in transport and communication costs. Of particular importance has been the development of various technologies for duplicating information, such as the typewriter used together with carbon copies, and more recently the photocopying and fax machines, and e-mail. Such duplicators have made the transmission of information along a managerial hierarchy and across space much faster and much cheaper. Other recent developments in information technologies, such computer-aided control systems, or online stock monitoring and ordering have also facilitated management over large distances.

Local population	Sectoral specialization[a]			Functional specialization in management against production[b]			
	1977	1987	1997	1950	1970	1980	1990
5,000,000–19,397,717	.377	.376	.374	+10.2%	+22.1%	+30.8%	+39.0%
1,500,000– 4,999,999	.366	.360	.362	+ 0.3%	+11.0%	+21.6%	+25.7%
500,000– 1,499,999	.397	.390	.382	−10.9%	− 7.8%	− 5.0%	− 2.1%
250,000– 499,999	.409	.389	.376	− 9.2%	− 9.5%	−10.9%	−14.2%
75,000– 249,999	.467	.442	.410	− 2.1%	− 7.9%	−12.7%	−20.7%
67– 75,000	.693	.683	.641	− 4.0%	−31.7%	−40.4%	−49.5%

Source: Duranton and Puga (2005) based on data from County Business Patterns (sectoral specialization) and Decennial Census of Population and Housing (functional specialization).

[a]Mean value for each population class of a Gini index comparing the local and national distributions of employment shares across 2-digit SIC manufacturing sectors. If s_h and $\bar{s}_h$ are respectively the local and national shares of employment in sector h, the Gini specialization index is $\frac{1}{2}\sum_h |s_h - \bar{s}_h|$. Its value is close to one if a city is fully specialized in a sector that is very small at the national level and is equal to zero if local employment is dispersed across sectors in the same way as national employment.

[b]Percentage difference from the national average in the number of executives and managers per production worker (occupied in precision production, fabrication, or assembly).

Table 1: The falling sectoral specialization and rising functional specialization of US cities

Duranton and Puga (2005) model the location decisions for firms with a multi-stage production process in a general equilibrium model of an urban system. They show that, as transport improvements and technological developments facilitate the spatial fragmentation of activities within the firm, firms search for the best separate locations for management and for production. For headquarters this means locations with a strong presence of business service providers (such as legal services or advertising agencies) that they can share with other headquarters. For production facilities, this means places with other similar plants. Consistent with this framework, Henderson and Ono (2008) find that, in choosing the location of newly-established stand-alone headquarters, US manufacturing firms trade off a greater variety of business service providers and a strong presence of other headquarters against proximity to the firm's production facilities. In the Duranton and Puga (2005) model, the combined location choices of many firms make each other worthwhile and change the nature of urban specialization. Headquarters end up concentrating in bigger cities to share service providers because business services tend to exhibit greater economies of agglomeration —the fact that they are less land-intensive pushes in the same direction. The ensuing increase in land prices prompts production establishments to relocate to smaller, more specialized, towns and cities. Duranton and Puga (2005) show that such a process has occurred in the United States since the 1950s. The pattern can be seen in table 1, taken from their paper. This classifies cities (Metropolitan Areas plus those counties not included in any Metropolitan Area) into 6 different size classes in terms of their population in 2000. Here we focus on the last four columns of the table. For each city size class, the numbers in these columns are constructed by calculating the ratio of executives and managers to production workers (occupied in precision production, fabrication, or assembly). They then list the percentage difference between this ratio and the corresponding ratio for the entire nation in 1950, 1970, 1980 and 1990. In 1950, the ratio of managers to production workers was similar across cities of different sizes. By 1980 differences across cities had increased substantially and a clear ranking by size had emerged: larger cities had become specialized in management functions whereas smaller cities had become specialized in production. This pattern became even more marked over the following decade. By 1990, cities with between 75,000 and 250,000 people had 20.7 percent fewer managers per production worker than the national average, whereas cities with between 1.5 and 5 million people had 25.7 percent more managers per production worker. Cities larger than 5 million people were 39 percent above the national average. A noteworthy feature of this process of functional specialization, evident in table 1, is that it does not emerge as a dichotomous difference between two types of cities, big and small. Instead the degree of specialization in management versus production changes gradually with city size. This is common to many other empirical results reviewed in this paper. There is no clear cutoff point distinguishing big and small cities, but numerous important gradients that are smooth functions of city sizes.

Several recent studies have found similar trends in other developed countries (see, e.g., Brunelle, 2013, for Canada and Bade, Bode, and Cutrini, 2015, for Germany). In effect, this implies an increasing specialization by functions and occupations instead of traditional sectoral divisions.

An important implication of the growing functional specialization of cities is that manufacturing has become less spatially concentrated whereas services have become more spatially concentrated. Holmes and Stevens (2004) compare the location of large manufacturing plants in the United States in 1947 and 1999 and show that industry has greatly dispersed away from its earlier concentration in big cities, particularly those in the Northeast and around the Great Lakes. According to Holmes and Stevens (2004), if we rank locations in the United States from the most to the least urbanized and look at the group of densest cities accounting for 25% of total US employment, they now only have 19% of manufacturing employment. Looking instead at the group of least dense urban and rural locations also accounting for 25% of total US employment, they now account for 30% of manufacturing employment.

Just as the main production facilities have left big cities, these cities have increasingly attracted business service sectors. Desmet and Rossi-Hansberg (2009) show that the process of geographic concentration that service employment has experienced over the last few decades is very similar to that experienced by manufacturing employment in 1900–1920. They argue that this is related to the diffusion of general purpose technologies in young industries. A new general purpose technology can only reach its full potential after much experimentation and many incremental innovations. The development of these incremental innovations is facilitated by spatial concentration and the agglomeration economies that concentration gives rise to. Subsequently, these smaller innovations diffuse and disperse to other areas. As a result, argue Desmet and Rossi-Hansberg (2009), in young industries if we relate growth in employment to its initial density, at low initial employment density levels, growth decreases with density (reflecting high growth potential through diffusion for locations with very small initial density). At intermediate initial density levels, growth then increases with density (reflecting the advantage of density for the development of new technologies through agglomeration economies). Finally, at high initial density levels, growth again decreases with density (reflecting congestion). Mature industries, in contrast, have standardized technologies that do not benefit much from agglomeration economies for further development. Employment growth in mature industries is thus negatively related to initial density.

Rather than distinguishing manufacturing from services, Michaels, Rauch, and Redding (2012) study manufacturing and services combined relative to agriculture. They show that this analysis can help explain a similar overall relationship between growth and initial density for population over a long horizon as that found by Desmet and Rossi-Hansberg (2009) for employment in new industries. In their framework, there is mean reversion in local agricultural employment: when a location experiences a shock to its level of agricultural employment, this eventually reverts towards a long-run level determined by permanent considerations such as climate and soil. In manufacturing and services, in contrast, permanent factors matter less, so that one does not observe mean reversion. At low population densities, agricultural employment is comparatively large, and mean reversion in agriculture generates a decreasing relationship between population growth and initial population density. At intermediate population densities, growth patterns are driven by the reallocation of employment across sectors over time: faster employment growth in manufacturing and services than in agriculture, combined with a concentration of manufacturing and services in areas with high initial density, generates an increasing relationship between population growth and initial population density. At high population densities, manufacturing and services employment is comparatively large, and the absence of mean reversion implies that population growth is mostly uncorrelated with initial population density.

3. City size and firm productivity

We have seen that big cities tend to attract different sectors and types of activities than small cities. However, differences between big and small cities are not just found at such aggregate levels. Even plants within the same sector and type of activity are systematically different between big and small cities. In particular, plants located in bigger cities exhibit significantly greater productivity than plants in smaller cities and rural areas. In a pioneering study, Sveikauskas (1975) regressed log output per worker in a cross-section of city-industries on log city population and found an elasticity of around 0.06. Given that capital is likely to be used more intensively in large cities, measuring productivity through output per worker as in this early study is likely to yield an overestimate of the relationship between city size and productivity. Thus, most recent studies use total factor productivity instead. Such studies find that a doubling of city size increases productivity by between 3 and 8 percent for a large range of city sizes (Rosenthal and Strange, 2004).

A crucial concern when studying the relationship between city size and productivity is that such a relationship is not necessarily causal. This is because city size and productivity can be simultaneously determined. Do larger cities lead to higher productivity or is there some underlying productivity advantage in some cities that then makes these cities grow to become larger? Following Ciccone and Hall (1996), a first strategy to tackle this issue is to instrument for the current size or density of an area. The usual instruments are historical population data for cities and characteristics that are thought to have affected the location of population in the past but that are mostly unrelated to productivity today. The logic behind these instruments is that there is substantial persistence in the spatial distribution of population (which ensures the relevance of the instruments), but the drivers of high productivity today greatly differ from those in the distant past (which helps satisfy the exclusion restriction). Most studies find that reverse causality and simultaneity are only minor issues in this context since estimates of the productivity advantages of larger cities are not substantially affected by instrumenting (Ciccone and Hall, 1996, Combes, Duranton, Gobillon, and Roux, 2010).

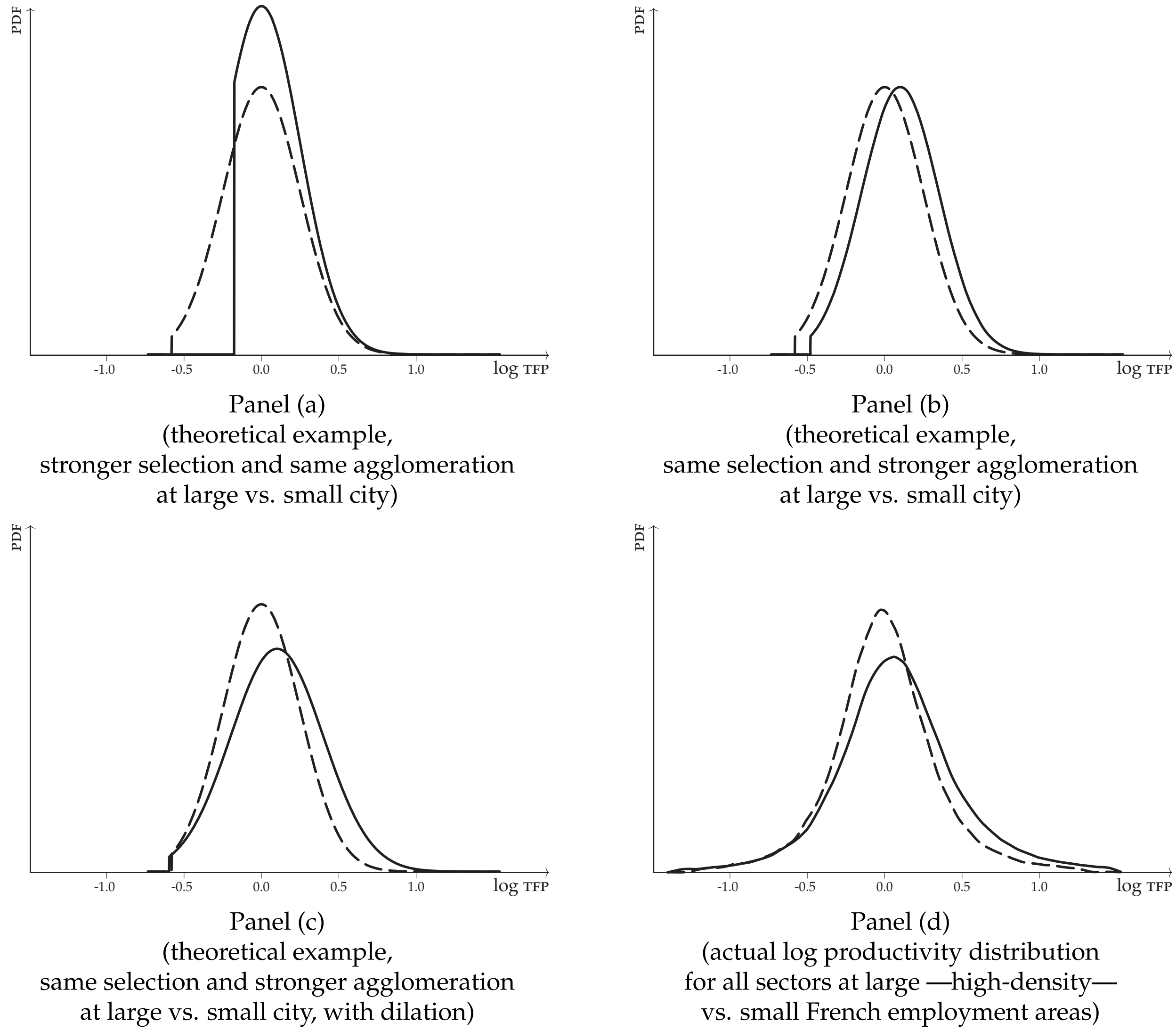
A second strategy to tackle simultaneity in the determination of city size and productivity is to use panel data. One can then include plant fixed effects when estimating plant-level productivity to capture any unobserved attributes that may have attracted more establishments to a given city (Henderson, 2003). While sound, this strategy has the practical problem that it relies on variation over time in both plant-level activity and measures of city scale. While there is often large variation over time at the plant level, there is much less variation at the city level, in particular in measures of city scale (a feature that the instrumentation strategy discussed above takes advantage of). Using lower frequency data to partly get around this then raises concerns about simultaneity in the determination of employment changes and productivity shocks.

A third strategy is to take advantage of some quasi-experimental setting. Greenstone, Hornbeck, and Moretti (2010) identify US counties that attracted large new plants (with investments above one million dollars) as well as runner-up counties that were being considered as an alternative location by the firm. The strategy relies on the fact that the winning and runner-up counties are likely to be very similar in every respect up until the investment takes place (which they show is indeed the case). Following the establishment of the large new plant, incumbent plants in winning counties experience a sharp increase in total factor productivity relative to incumbent plants in runner-up counties. The effect is stronger for incumbent plants that share similar labor and technology pools with the new plant.

All of this points to the existence of a causal relationship in which larger cities lead to higher productivity. However, that still does not guarantee that this is a consequence of agglomeration economies, whereby individual firms see their productivity rise due to the benefits of being located close to many other firms. It could also be the case that the large number of firms in bigger cities makes competition tougher, as in Melitz and Ottaviano (2008), reducing markups and inducing less productive firms to exit. In this case, higher average productivity in larger cities could result from firm selection eliminating the least productive firms.

Combes, Duranton, Gobillon, Puga, and Roux (2012*b*) develop a framework to distinguish between agglomeration and firm selection. They nest a generalized version of the firm selection model of Melitz and Ottaviano (2008) and a simple model of agglomeration in the spirit of Fujita and Ogawa (1982). This nested model enables them to parameterize the relative importance of agglomeration and selection. The main prediction of their model is that, while selection and agglomeration effects both make average firm log productivity higher in larger cities, they have different predictions for how the shape of the log productivity distribution varies with city size. This can be seen in panels (a), (b), and (c) of figure 1, taken from figure 3 in Combes, Duranton, Gobillon, Puga, and Roux (2012*b*). The three panels illustrate their model by plotting the distributions of firm log productivity in a city with a large population (continuous line) and in a city with a small population (dashed line) under different assumptions of what varies across cities of different sizes. Panel (a) is plotted under the assumption that competition between firms is fiercer, and thus markups lower, in big than in small cities, leaving less productive firms who would have survived in a small city unable to make it in a big city. Such stronger selection effects in big cities, by excluding the least productive firms, should make the distribution of firm log productivities in big cities left-truncated relative to the distribution in small cities, as shown in panel (a). In panel (b) it is instead assumed that selection effects are no different in big and in small cities, but the presence of more firms makes productivity-enhancing agglomeration economies stronger in big than in small cities. Stronger agglomeration effects, by making all firms more productive, should lead instead to a rightwards shift of the distribution of firm log productivities in big cities relative to small cities, as shown in panel (b). That second panel is drawn under the simplifying assumption that all firms benefit equally from agglomeration economies. If firms that are more productive are also better at reaping the benefits of agglomeration, then agglomeration should lead not only to a rightwards shift but also to an increased dilation of the distribution of firm log productivities in larger cities, as shown in panel (c).

A comparison of the entire distributions of firm log total factor productivities in big cities and in small cities is thus informative of how the extent of firm selection and the strength of agglomeration effects vary with city size. Combes, Duranton, Gobillon, Puga, and Roux (2012*b*) develop a quantile approach that allows estimating a relative change in left truncation, shift, and dilation between two distributions. Applying this approach to establishment-level data for France, they conclude that productivity differences across urban areas in France are mostly explained by agglomeration. They compare locations with above median employment density against those with below-median density —results are almost identical when comparing the log productivity distribution of large cities with population above 200,000 with that of cities smaller than that. The distribution of firm log productivity in areas with above-median density is shifted to the right



Panel (a)
(theoretical example,
stronger selection and same agglomeration
at large vs. small city)

Panel (b)
(theoretical example,
same selection and stronger agglomeration
at large vs. small city)

Panel (c)
(theoretical example,
same selection and stronger agglomeration
at large vs. small city, with dilation)

Panel (d)
(actual log productivity distribution
for all sectors at large —high-density—
vs. small French employment areas)

Source: Combes, Duranton, Gobillon, Puga, and Roux (2012*b*).

Figure 1: Comparisons of log productivity distributions in large (solid) and small cities (dashed) under differences in selection and/or agglomeration

and dilated relative to areas below median density. On the other hand, they find no difference between denser and less dense areas in terms of left truncation of the log productivity distribution, indicating that firm selection is of similar importance in cities of different sizes. Panel (d) in figure 1 plots the actual empirical distributions of log productivity in French employment areas with above-median employment density (continuous line) and below-median employment density (dashed line). We can see how similar this is to the theoretical benchmark of panel (c), plotted under the assumption of same selection and stronger agglomeration (with more productive firms reaping greater benefits from this) at large vs. small cities. The methodology of Combes, Duranton, Gobillon, Puga, and Roux (2012*b*) allows for a more formal comparison of the distrubutions, and their results show that firms in denser areas are on average about 9.7 percent more productive than in less dense areas. However, the productivity boost of larger cities is greater for more productive firms, so the productivity gain is 14.4 percent for firms at the top quartile and only of 4.8 percent for firms at the bottom quartile.

The estimates of Combes, Duranton, Gobillon, Puga, and Roux (2012*b*) imply an elasticity of productivity with respect to employment density of 0.032, which is in line with estimates of agglomeration economies in the literature. While there is by now broad agreement on the existence and approximate magnitude of agglomeration economies at the urban level, urban economists have been far less successful at distinguishing between the possible sources of these agglomeration economies. On the theoretical front, there is a large literature that develops three broad classes of mechanisms to explain the existence of urban agglomeration economies (this classification follows Duranton and Puga, 2004, who cover the theoretical literature in detail). First, a larger market allows for a more efficient *sharing* of local infrastructure and facilities, a variety of intermediate input suppliers, or a pool of workers with similar skills. Second, a larger market also allows for a better *matching* between employers and employees, or buyers and suppliers. This better matching can take the form of improved chances of finding a suitable match, a higher quality of matches, or a combination of both. Finally, a larger market can also facilitate *learning*, for instance by facilitating the transmission and accumulation of skills or by promoting the development and widespread adoption of new technologies and business practices.

Identifying these possible mechanisms and assessing their relative importance is difficult because they all share the prediction that productivity increases with the scale of an activity at a location. This 'Marshallian equivalence' (Duranton and Puga, 2004) makes it very difficult to distinguish the main causes of the productivity advantages of big cities.

One common empirical strategy, used among others by Audretsch and Feldman (1996), Rosenthal and Strange (2001), and Overman and Puga (2010), is to measure the geographical concentration of different sectors and regress this on proxies for different mechanisms. This amounts to checking whether the most concentrated sectors are, for instance, those for which sharing intermediate inputs is relatively important, or those that have a greater potential for labour pooling, or those for which innovation and knowledge spillovers matter more as a way to compare the relative importance of these mechanisms for agglomeration. These regressions yield evidence for the various motives for agglomeration. Interestingly, their relative importance varies with the scale of the analysis, so that industries for which transport and access to inputs are important tend to cluster at aggregate geographical levels whereas industries for which innovation is important tend to cluster at much finer geographical levels (Rosenthal and Strange, 2001). This suggests a different rate of spatial decay for different types of spillovers. For instance, knowledge spillovers likely operate only over very close areas, such as city neighbourhoods, where frequent face-to-face interactions are possible. Labour market polling likely operates at the level of local labour markets, which correspond to economic definitions of cites. Input sharing instead likely operates over quite large regions. Also, reliance on certain inputs leads to spatial concentration only if those inputs are themselves spatially concentrated: sectors that purchase lots of intermediate inputs are not necessarily more concentrated, whereas those that buy lots of inputs found only in few places are significantly more concentrated (Overman and Puga, 2010).

Instead of assessing the relative importance of possible agglomeration mechanisms by looking at which industries are more and less spatially concentrated, Ellison, Glaeser, and Kerr (2010) study co-agglomeration patterns. The underlying idea is that, while plants in any given industry are similar in many dimensions, when we look across industries plants are similar in some dimensions and not in others. For example, some industry pairs employ similar types of workers but purchase different intermediate input mixes. Others hire different types of workers but purchase similar intermediate input mixes. Hence, by studying which similarities across industries help to predict better which industry pairs are co-agglomerated, one can gain insight into the relative importance of different motives for agglomeration. Ellison, Glaeser, and Kerr (2010) find that sectors buying similar intermediates tend to co-agglomerate the most, followed by sectors that employ similar workers. To address reverse causality (i.e., the possibility that certain industry pairs end up buying intensively from each other precisely because they are agglomerated), they use UK measures to instrument for US industry characteristics.

The two strategies described above run a horse race between multiple potential sources of agglomeration economies. Other papers try instead to isolate one mechanism in particular. For instance, Holmes (1999) studies the importance of access to local input suppliers. There are two reasons why this may matter. First, as in Abdel-Rahman and Fujita (1990), compared to a firm in a location with a sparse supplier network, a firm in a location with a dense supplier network has access to a wider variety of intermediate inputs. Second, as discussed by Stigler (1951), a firm in a location with a relatively dense supplier network can outsource the production of more inputs instead of having to produce them in-house, which can help reduce costs. Holmes (1999) combines detailed plant-level data for the United States with spatial data on input purchases. He begins by showing that firms in industries that are more spatially concentrated purchase more inputs locally instead of sourcing them far away. He then regresses the intensity of local input purchases on same-industry employment in the plant's own county and in other counties within fifty miles (with both measures differenced from the industry mean). Holmes (1999) finds that the purchased-inputs intensity of a plant increases with the level of employment of nearby plants in the same industry.

Jaffe, Trajtenberg, and Henderson (1993) study instead knowledge spillovers. While knowledge flows are generally hard to track and measure, they develop an ingenious strategy to do so using patent data. Patents contain among other information the name and address of the inventors as well as citations to other patents that are deemed relevant prior knowledge. Jaffe, Trajtenberg, and Henderson (1993) show that patents tend to cite disproportionately patents by inventors who are located nearby. Of course, this could reflect that patents tend to cite mainly patents from the same or similar industries and that industries can be spatially concentrated for different reasons. To get around this issue, Jaffe, Trajtenberg, and Henderson (1993) construct a set of control patents, similar in terms of technology and timing to the patents in their citation data. They show that a patent is much more likely to have an inventor at the same location as another patent it cites than it is to have an inventor at the same location as a matched patent from this control group. Thompson and Fox-Kean (2005) argue that the results of Jaffe, Trajtenberg, and Henderson (1993) are sensitive to the procedure used to match control patents. Jaffe, Trajtenberg, and Henderson (1993) match patents from the same three-digit patent class (of about 450 possible classes) and find strong evidence of localized spillovers. Thompson and Fox-Kean (2005) match patents from the same patent subclass (of about 150,000 possible subclasses) and find much weaker evidence. Murata, Nakajima, Okamoto, and Tamura (2014) argue that since patent subclasses are so much narrower, they are also likely to be localized at a finer geographical level. For this reason, instead of studying the localization of patent citations within metropolitan area or state boundaries, as Jaffe, Trajtenberg, and Henderson (1993) and Thompson and Fox-Kean (2005) do, Murata, Nakajima, Okamoto, and Tamura (2014) use a point-based measure drawing on Duranton and Overman (2005) that studies pairwise distances between inventors both for patent pairs reflecting actual citations and for patent pairs used as controls. With this methodology, they find support for localized spillovers even when using the same detailed patent subclass level as Thompson and Fox-Kean (2005).

One may wonder whether improvements in communications technologies have reduced the importance of proximity for innovation or whether, on the contrary, they have contributed to make innovation so rapid and complex that constant face-to-face interactions are essential to stay at the frontier. Sonn and Storper (2008) suggest it is the latter and find that the importance of geographical proximity for knowledge flows (measured through patent citations) has increased over time.

The importance of proximity in patent citations is in part related to its importance in research collaborations. Thus, inventors who change jobs very often do not change location and keep citing disproportionately former co-inventors (now at a different employer but still in their same location) or other people within their larger collaboration network (Breschi and Lissoni, 2009). When inventors do relocate, their ties with inventors with whom they were previously co-located tend to persist (Agrawal, Cockburn, and McHale, 2006). And other types of connections are also important, so that geographical proximity matters less when inventors are connected by social or ethnic ties (Agrawal, Kapur, and McHale, 2008). In fact, it has been argued that, although most innovations arise in big cities (Feldman and Audretsch, 1999), peripheral regions can promote innovation by fostering connections beyond their immediate proximity (Rodríguez-Pose and Fitjar, 2013). Geographical proximity can thus be seen as one of several factors helping establish connections among inventors, with these connections in turn facilitating knowledge flows.

Of course not all innovations are patented, and technical progress is related not just to the initial development but also to the speed of adoption of new technologies. Another strand of literature focusing on knowledge spillovers studies how the local environment affects the probability of adoption of advanced technologies. Kelley and Helper (1999) study the adoption of a novel technology, computer numerically controlled machines, by individual establishments in a variety of machine-making industries (ranging from heating equipment and plumbing fixtures to guided missiles and aircraft). They find that the probability of adoption by an establishment greatly increases when the establishment is in an area that has a diversity of sectors as opposed to a very narrow specialization, and if the establishment's manager has learned about the technology from multiple sources, including public presentations, visits by sales representatives, and conversations about the technology with managers at other firms. No (2003), using data on the adoption of advanced manufacturing technologies in Canada, finds that adoption is more likely in locations with more prior adopters. The effect is stronger when these prior adopters use similar technologies as the establishment being considered but do not compete in the exact same sector —so that there are fewer reasons for managers of prior adopters to hide information about these novel technologies.

4. Worker skills and productivity across cities

Just as bigger cities tend to host more productive establishments, they are also increasingly home to a disproportionate share of high-skilled and particularly productive workers. Looking first at skills, urban areas tend to have better-educated people than rural areas. And comparing different urban areas with each other, bigger cities also feature higher proportions of workers with greater education and more skilled occupations (Berry and Glaeser, 2005, Moretti, 2012, Davis and Dingel, 2013). These sort of differences have grown over time. Bigger cities already had more educated workers early on, but then cities with greater initial proportions of highly-educated workers have grown more over time (Berry and Glaeser, 2005, Moretti, 2012).

There is debate about whether big cities, in addition to having more workers at the high end of the skill distribution, also have more workers at the low end, so that the overall distribution of observable skills has greater variance in bigger cities. Eeckhout, Pinheiro, and Schmidheiny (2014) suggest this is the case in the United States. They attribute this feature to extreme complementarities in production (say, doctors benefiting more from the presence of hospital janitors than that or nurses). Another possible explanation is complementarities in consumption instead of production. According to Moretti (2012) for every high-skilled job in a high-tech sector in a city, five additional jobs are created outside the high-tech sector. Some of these are high-skilled (lawyers, teachers) and others low-skilled (waiters, hairdressers, nannies). Finally, it could also be the case that low-skilled workers are attracted to big cities by amenities or public goods that are particularly prevalent there, including public housing and public transport. Related to this last argument is the idea that foreign-born workers are the majority of the group of workers with the lowest skills in the United States and the foreign-born may find it easier to settle in bigger cities. Davis and Dingel (2013) support this argument by showing that workers with less than high school education were overrepresented in big cities in 2000, when foreign-born workers accounted for 80 percent of this group, but not in 1980, when foreign-born workers only accounted for one-third of this group.

Workers in bigger cities not only have higher education on average, they also have higher nominal earnings even after discounting the effect of their greater education. Higher nominal earnings do not necessarily mean that workers are better off in bigger cities, since they also have to pay higher prices for housing and also experience other disadvantages from density, such as greater congestion.

While there is much empirical work quantifying the advantages of bigger cities in terms of higher productivity and earnings, much less is known about how urban costs increase with city size. Bettencourt, Lobo, Helbing, Kühnert, and West (2007) estimate log-linear relationships between several magnitudes and city sizes. They find positive coefficients for total output and total wages, which is equivalent to the positive elasticity of productivity and individual earnings with respect to city size identified in the literature reviewed above (although the evidence in Bettencourt, Lobo, Helbing, Kühnert, and West, 2007, is descriptive and does not attempt to establish a causal relationship with city size). Interestingly, they find negative coefficients for shared infrastructure (total electrical cable length and total road surface), implying that such infrastructure scales less than proportionately with city size. This could partly reflect economies of scale in infrastructure provision (for instance, the cost per household of building an electricity network or a water network is lower when supplying densely populated areas). However, it could also partly reflect congestion (for instance, as cities grow over time, existing buildings place constraints on road expansion in city centres). To the extent that city dwellers bear the cost of shared infrastructure, economies of scale in infrastructure provision imply that urban costs per person fall with city size whereas congestion instead implies that urban costs per person increase with city size.

Rather than trying to disentangle different components of urban costs, a fruitful approach is to summarize them through housing costs at the city centre. Out in the suburbs, urban costs are a combination of housing costs, commuting costs, and possibly other costs related to reduced accessibility. However, at the centre of any city the cost of commuting can be regarded as negligible or at least comparable to other city centres. Since mobility within the city will tend to equalize total urban costs, housing costs at the centre will be comparable to the various combinations of urban costs experienced at other points in the city. On the basis of this idea, Combes, Duranton, and Gobillon (2012*a*) develop a methodology to estimate how urban costs vary with city size and apply it to French data. They use information about the location of parcels in each city and other parcel characteristics recorded in land parcel transactions to estimate unit land prices at the centre of each city. They then regress the log of these estimated prices at the centre of each city on log city population to obtain an estimate of the elasticity of unit land prices at the centre of each city with respect to city population. Multiplying this by the share of land in housing and then by the share of housing in expenditure, yields an elasticity of urban crowding costs with respect to population of 0.041. This is similar but slightly higher than most estimates of the elasticities of productivity or wages with respect to city size. If we think of the overall benefits and costs of living in a city as the difference between two exponential functions of population, one capturing agglomeration economies and another one capturing urban crowding costs then having a somewhat higher exponent on the crowding term is necessary for urban population not to be completely absorbed by a single city with unbounded growth (Duranton and Puga, 2014).

Higher costs of living explain why workers do not flock to bigger cities. However, higher earnings in bigger cities still indicate that bigger cities provide productive advantages. If firms in tradable sectors did not find workers in bigger cities more productive, they would locate in smaller cities instead. While not all firms are in tradable sectors, as Moretti (2011) notes, "as long as there are some firms producing traded goods in every city and workers can move between the tradable and non-tradable sector, average productivity has to be higher in cities where nominal wages are higher."

As in the case of firms, a key question is whether workers in bigger cities become more productive because they locate there, of if instead workers who are more productive to start with for other reasons end up sorting into bigger cities. A first strategy to deal with this issue is to control for education and occupational skills, and this still leaves a substantial earnings premium associated with working in bigger cities. However, it is possible that workers are different not only along dimensions we observe, such as education, but also along dimensions we do not observe, such as ability. A strategy to address the issue of workers sorting across cities on unobservables is to use panel data following workers over time and across cities and introduce worker fixed-effects to account for any time-invariant individual worker characteristics (Glaeser and Maré, 2001, Combes, Duranton, and Gobillon, 2008, De la Roca and Puga, 2017), an approach to address the issue of workers sorting across cities on unobservables is to introduce worker fixed-effects. This approach yields and elasticity of the earnings premium with respect to city size of 0.025–0.03.

We have seen that big cities have more highly-educated workers and, even after accounting for those differences, workers earn more in bigger cities. However, within broad occupation or education groups, it is not the case that the most able workers are disproportionately in big cities. The same conclusion is reached regardless of whether we measure ability through cognitive test results (Bacolod, Blum, and Strange, 2009), individual fixed-effects in a wage regression (De la Roca and Puga, 2017), measures of ability derived from a finite-mixture model in a structural estimation setting (Baum-Snow and Pavan, 2012), or individual residuals from a spatial equilibrium condition (Eeckhout, Pinheiro, and Schmidheiny, 2014). This is all the more surprising since the benefits that bigger cities provide in terms of higher earnings are greater for more able workers (De la Roca and Puga, 2017, Baum-Snow and Pavan, 2012).

One reason for little sorting on ability across cities within broad occupation or education groups is that mobility is limited. Even in the United States, 56% of people live at age 40 in the same city or town where they were at age 14 (De la Roca, Ottaviano, and Puga, 2014). Another reason, suggested by De la Roca, Ottaviano, and Puga (2014), is that individuals only have an imperfect assessment of their own ability, especially when they are young. Workers whose self-confidence at an early stage of their career is not aligned with their ability may make location decisions they would not have made if they had known their actual ability to start with. By the time they learn enough about their actual ability, those early decisions have had a lasting impact, reducing their incentives to move and affecting their lifetime earnings.

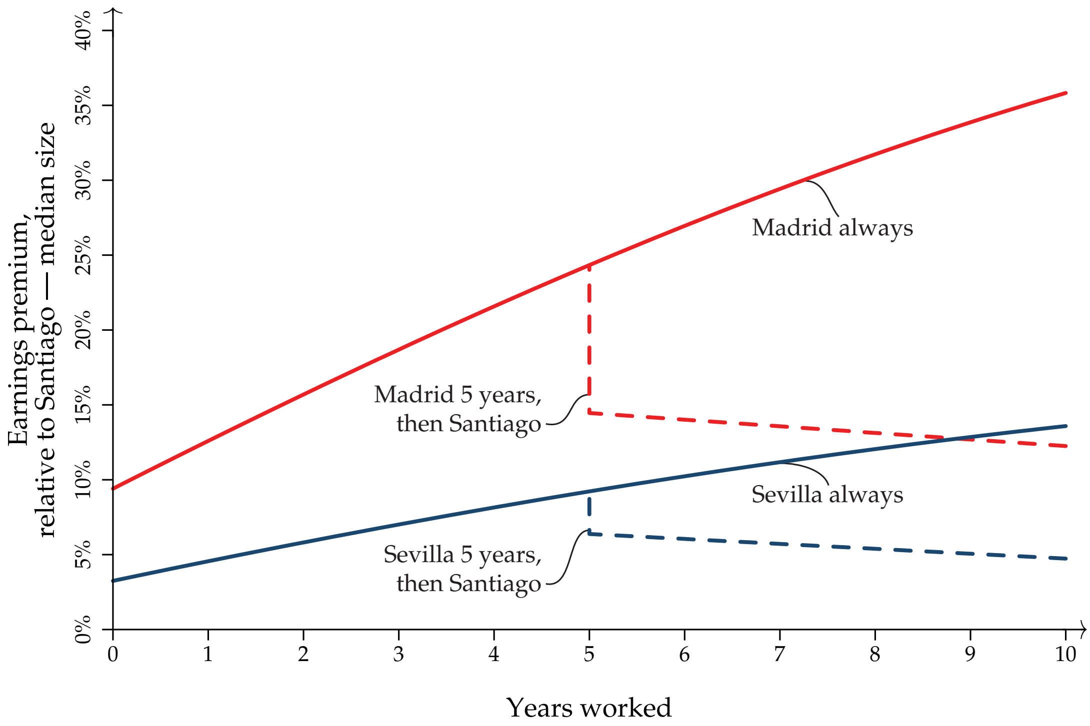
De la Roca, Ottaviano, and Puga (2014) test and find support for this idea using panel data for individual US workers that includes location, personal and job characteristics, measures of ability and measures of self-confidence. They find that self-confidence is more important than actual ability for the location decisions of young workers. For older workers, ability plays a stronger role in determining location, but the lasting impact of their earlier choices limits the scope for relocation. However, workers who seriously underestimate their own ability may nevertheless relocate from a small to a big city once their labour market experience provides them with better information of their true capabilities. Relocations from big to small cities appear to be driven instead by lack of success in the big city rather than by corrections to flawed self-assessment.

5. Life-cycle patterns of location across cities

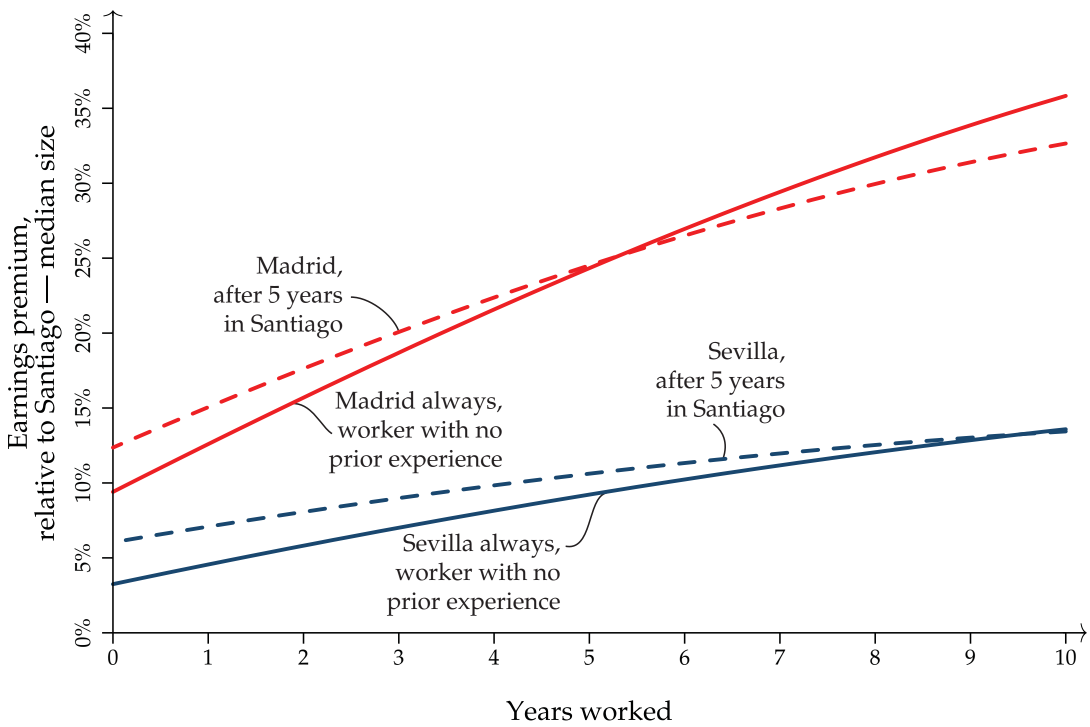
All of the above may give the impression that bigger cities hold an increasingly advantageous position relative to smaller cities. And yet, this is not the case. First, because big cities not only feature higher firm productivity and worker earnings, they also have more congestion and substantially higher land and house prices. Furthermore, because as cities have become more different, they have also become more interdependent in many ways. Thus, it would be simplistic to think of urban systems as made of 'winner' and 'loser' cities. Instead, important dynamic patterns have emerged linking cities more closely to each other.

Workers are often willing to pay the higher living cost of bigger cities not just to get higher wages today, but in the hope of learning and acquiring skills that, depending on their own characteristics and luck, they may exploit there or somewhere else. De la Roca and Puga (2017) examine three reasons why firms may be willing to pay more to workers in bigger cities. First, there may be some static advantages associated with bigger cities. Second, bigger cities may allow workers to accumulate more valuable experience. Third, workers who are inherently more productive may choose to locate in bigger cities.

De la Roca and Puga (2017) find that there are substantial static and dynamic advantages from working in bigger cities. For the average worker, about one-half of these gains are static and tied to currently working in a bigger city. However, about another half accrues over time as workers accumulate more valuable experience in bigger cities. This can be seen in figure 2, which combines two figures from their paper. Panel (a) plots the evolution of earnings for workers in cities of different sizes, calculated based on regression coefficients estimated using matched individual social security, census and income tax data for Spanish workers. Their estimations allow the value of worker experience to vary depending both on where the experience was acquired and where it is being used. The top solid line corresponds to the earnings profile over ten years of an individual with no prior experience who works in Madrid (the largest city) over the entire period relative to the earnings of a worker with identical characteristics (both observable and time-invariant unobservable) who instead works in Santiago de Compostela (the median-sized city). Already upon starting to work, the individual in Madrid earns almost 10% more than the worker in Santiago. In addition to this, experience acquired in Madrid appears to be substantially more valuable than experience acquired in small cities and this gets reflected in the earnings gap



Panel (a) The learning benefits of bigger cities for difference length of stays



Panel (b) Comparing workers with and without prior experience

Source: De la Roca and Puga (2017) based on data from Muestra Contínua de Vidas Laborales.

Figure 2: Earnings profiles relative to median-sized city

between the worker in madrid and the one in Santiago widening considerably over time, reaching over 35% after 10 years. The top dashed line in panel (a) of figure 2 shows that the additional value of the experience acquired in Madrid remains even if the worker moves away. The line shows the estimated relative earnings profile for an individual who, after five years of working in Madrid, moves to Santiago. Up until year five, his relative earnings profile is the same as that of a worker who always works in Madrid. At that point, he relocates to Santiago, and his relative earnings drop as a result. Note that the magnitude of this drop is very similar to the initial gap, implying that upon relocating to Santiago the worker is able to take with him roughly all of the additional additional value of experience acquired during the five years of working in Madrid. From year 5 onwards the line has a slight negative slope, showing a slow depreciation of the experience acquired in Madrid after moving away. The bottom solid and dashed lines in panel (a) of figure 2 show similar profiles, even if smaller in magnitude, for a worker locating in Sevilla (the fourth largest city) relative to the earnings of a worker in Santiago. This suggests that an important part of the benefits of bigger cities are related with workers learning by working there, with this benefits becoming embedded in their human capital and being highly portable across locations.

Panel (b) of figure 2 shows that bigger cities, in addition to providing workers with more valuable experience, also reward more experience acquired elsewhere. The solid lines are the same as in panel (a), drawn for a worker with no prior experience. The dashed lines are drawn instead for a worker with 5 years of prior experience acquired in Santiago. These 5 years of prior experience imply a larger immediate gain upon moving to Madrid than in the absence of prior experience, because employers in Madrid reward prior experience more than employers in Santiago. However, thereafter the dashed profiles for a worker with prior experience are flatter than the solid profiles for a worker with prior experience, because big city experience has a greater impact on less experienced workers.

Just as there are important life-cycle patterns across cities for workers, these also exist for firms. Many new firms and products start up in cities that are bigger, more educated and more diverse but eventually relocate, making up space for the next generation of new firms and products. Duranton and Puga (2001) propose a model where the benefits from learning in one city can be exploited in another. In this model, incurring a fixed cost allows an entrepreneur to introduce a new product. However, initially, this will only be a prototype. Developing the final standardized product starting from this prototype requires experimentation and learning. In particular, entrepreneurs can choose between many production processes, each associated with a different set of inputs. The ideal production process, which differs across entrepreneurs, is initially unknown. An entrepreneur can try to discover her ideal production process by sampling at most one production process each period and using it for prototype production. As soon as an entrepreneur samples her ideal production process, she recognizes this and can begin mass-production. To capture ongoing firm entry and the fact that learning is never exhausted, a proportion of firms is assumed to randomly exit every period. In addition, relocating production across cities is costly. The model assumes that there are agglomeration economies, so that the cost of using a given production process diminishes as more local firms use the same type of process. At the same time, urban crowding places a limit on city size and consequently on how many processes can be widely used in a city.

In the Duranton and Puga (2001) model, there is an equilibrium where both diversified and specialized cities arise endogenously, provided that learning is important and that establishment relocation costs are neither too high nor too low. This reconciles the needs for diversity and specialization along the life-cycle of firms. Entrepreneurs develop new products in cities with a diversified production structure. It allows them to sample easily and discover their ideal set of inputs. After discovering this ideal set of inputs, entrepreneurs are no longer interested in urban diversity. Because input producers in different sectors do not benefit from each other directly, industrial diversity makes cities more costly. As a result, entrepreneurs who have discovered their ideal set of inputs move away from a diversified city to a specialized cities so that they can benefit from agglomeration effects in the production of those inputs. Moving costs cannot be too high for relocation to occur after learning, nor so low that an entrepreneur can easily learn by constantly relocating. Further, the gains from learning need to be high enough to justify the foregone static agglomeration economies in the early phases. In this sense, we can think of diversified cities as 'nursery cities' where learning takes place and specialized cities as the places where the production of mature goods occurs.

The nursery cities model of Duranton and Puga (2001) proposes a theory of how innovation takes place and diffuses in space, while also matching observed patterns of firm relocations and a number of other facts about cities. We have seen above that big cities tend to host more management and business service activities than other areas. Bigger and more diverse cities also disproportionately give rise to new and innovative activities whereas more mature activities tend lo locate in smaller and more specialized areas. Jacobs (1969) already stressed the advantages of large and diverse cities for innovation. To study this issue empirically, Feldman and Audretsch (1999) collect information on nearly 4,000 product innovations that were made in the United States in 1982 for which it is possible to identify the location of the establishment where the innovation took place. They find that 96% of all new product innovations in the United States took place in metropolitan areas, although these account for only 30% of the population. Big cities are thus the places where innovations occur. They then look within metropolitan areas, to study whether the main force driving innovation is pure city size, urban diversity, or urban specialization. For this purpose, they group together four-digit industries sharing a common scientific base and show that industries sharing a common scientific base tend to cluster together. Finally, they perform regressions of innovations in each industry-city pair on an index of relative specialization in the sector, an index of diversity across industries with a common science base, city size, and other controls. They find that own-industry specialization 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.

Another implication of the model of Duranton and Puga (2001) relates to location and relocation patterns. Studies of firm location confirm that new production establishments choose to locate disproportionately in more diverse cities (Lainé and Rieu, 1999, Holl, 2004). Duranton and Puga (2001) study patterns of establishment relocations across cities. They study every establishment relocation that took place in France between 1993 and 1996 (episodes in which the closure of an establishment is followed by the opening in a different location of an establishment owned by the same firm and performing the same range of activities). Their results are presented in table 2, reproduced from their paper. The geographical origin and destination of relocating establishments is identified at the level of employment areas (zones d'emplois). The diversity of the origin area is measured by the inverse of a Herfindahl index of the sectoral concentration of local employment. The specialization of the destination area is measured by the share of the establishment's sector in local manufacturing and business-service employment. Looking at the bottom row of table 2, which combines relocations in all sectors, we can see the 4.7 percent of all French establishments relocated during the four-year period. The key finding is that 72 percent of these relocations followed the pattern predicted by their model: they were from an area with above median diversity to an area with above median specialization in the corresponding sector. In their model, more

	Percentage of relocations from diversified to specialized areas[a]	Relocations as a percentage of the stock[b]	Geographic concentration[c]
R&D	93.0	8.1	0.023
Pharmaceuticals and cosmetics	88.3	6.4	0.020
IT and consultancy services	82.1	7.3	0.030
Business services	75.8	5.0	0.015
Printing and publishing	73.3	5.4	0.026
Aerospace, rail and naval equipment	71.6	3.3	0.026
Electrical and electronic equipment	69.1	4.2	0.011
Motor vehicles	62.5	2.7	0.020
Electrical and electronic components	60.9	5.9	0.007
Textiles	46.4	2.5	0.024
Chemical, rubber and plastic products	38.3	3.9	0.009
Metal products and machinery	37.6	3.2	0.005
Clothing and leather	36.3	3.4	0.013
Food and beverages	34.6	0.8	0.007
Furniture and fixtures	32.6	2.7	0.008
Wood, lumber, pulp and paper	30.6	1.7	0.009
Primary metals	30.0	2.5	0.009
Non-metallic mineral products	27.3	2.0	0.012
Aggregate	72.0	4.7	

Source: Duranton and Puga (2001) based on data from SIRENE and ESE.

[a]Percentage of all establishments relocating across employment areas that move from an area with above median diversity to an area with above median specialization.

[b]Establishment relocations across employment areas as a percentage of the average number of establishments.

[c]Ellison and Glaeser (1997) geographic concentration index.

Table 2: Establishment relocations across French employment areas 1993–1996

innovative and agglomerated sectors are likely to benefit the most from the advantages that diversity and specialization offer at different stages of the product-cycle. Consistent with this, when we look at different sectors in the table, we see that the pattern of relocating from diversified to specialized areas is is particularly strong for firms for which innovating and being close to similar firms are particularly important (such as electronics producers). It is worth noting that when such relocations take place, establishments tend to move to locations that are more specialized but still easily accessible from their previous location. Also, there are cases where exploiting the advantages of diversity early in a product's life cycle and the advantages of specialization later on does not involve relocation of a production facility but relocation of production across two facilities. In this respect, Fujita and Ishii (1998) study the location of innovative activities of nine major Japanese electronic firms and find similar results regarding the importance of urban diversity. They show that the 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.

The model of Duranton and Puga (2001) suggests that, even if innovation and learning concentrate in a few large and diverse cities, this does not imply that smaller and more specialized cities will disappear. Instead, the diffusion of innovations to exploit them in small specialized cities frees up large and diverse cities to concentrate in continuously feeding the growth process with new ideas. As Duranton and Puga (2001) put it, for manufacturing and services, unlike for agriculture, 'sowing' and 'reaping' can take place in different locations.

6. Summary and conclusions

The location of big cities has become less and less tied to places with certain natural advantages. Historically, when economies were almost entirely agricultural and trade was limited, proximity to fertile land was essential. As trade increased, access to oceans and navigable waterways gained importance. Evidence of this can be seen in the change in city locations in England after the collapse of the Roman Empire (Michaels and Rauch, 2015). Later, with industrialization, sources of energy and natural resources attracted the growing manufacturing sector (especially as the transition from water power to steam power allowed the urbanization of industry, Rosenberg and Trajtenberg, 2004). The tertiarization of advanced economies together with falling transport costs has to a large extent freed the location of big cities from such constraints.

Natural advantages now matter more for their amenity value (Glaeser, Kolko, and Saiz, 2001, Rappaport, 2007). Production instead is greatly influenced by agglomeration economies, that is the productive advantages that firms and workers derive from proximity to other firms and workers. Firms and workers want to be where other firms and workers are. This may be in places where they always have been —hence the great persistence of city locations and relative city sizes that allows land fertility, historical transportation routes and historical population as instruments for current city sizes (Ciccone and Hall, 1996, Combes, Duranton, Gobillon, and Roux, 2010, De la Roca and Puga, 2017). But it could also be in any other place where many firms and workers choose to locate together. This generates great potential for instability and change.

The relocation of entire cities requires great coordination that is difficult to achieve. However, shifts in the activities, sectors, and occupations prevalent in each city can happen more gradually and easily. Over the last few decades such changes in the distribution of firms and workers across cities in developed countries have changed the face of urban systems.

Cities used to differ in size and in sectoral specialization but were otherwise more similar to each other than today. Since then, as technological developments and transport improvements have facilitated the spatial fragmentation of activities within the firm, headquarters, management activities, and business service provision have tended to concentrate in larger cities. Actual production has instead concentrated in smaller cities. In effect, this implies an increasing specialization by functions and occupations instead of traditional sectoral divisions.

An important implication of this growing functional specialization of cities is that manufacturing has become less spatially concentrated whereas services have become more spatially concentrated. This process of geographic concentration that service employment has experienced over the last few decades is very similar to that experienced by manufacturing employment in the first two decades on the twentieth century.

Big cities do not just attract different sectors and types of activities than small cities. Even plants within the same sector and type of activity are systematically different between big and small cities. In particular, plants located in bigger cities exhibit significantly greater productivity than plants in smaller cities and rural areas. For instance, within France, locating in Lyon (population 1.5 million) instead of Nantes (600,000) increases plant-level productivity by an average of 5 percent. Being in Paris (population 10 million) instead of Lyon increases productivity by another 10 percent on average. Finding productivity benefits of locating in bigger cities of such magnitudes is typical. Furthermore, the use of instrumental variables, panel data with plant fixed effects, or quasi-experimental variation all support a causal relationship, whereby a greater spatial concentration of firms and workers leads to higher productivity. And examining the entire distribution of plant productivities instead of just its mean indicates that this relationship does not arise from stronger competition in bigger cities acting as a selection mechanism but because individual firms see their productivity rise due to the benefits of being located close to many other firms.

There is a large literature that develops three broad classes of mechanisms to explain the existence of such productive advantages, known as agglomeration economies. First, a larger market allows for a more efficient sharing of local infrastructure and facilities, a variety of intermediate input suppliers, or a pool of workers with similar skills. Second, a larger market also allows for a better matching between employers and employees, or buyers and suppliers. This better matching can take the form of improved chances of finding a suitable match, a higher quality of matches, or a combination of both. Finally, a larger market can also facilitate learning, for instance by facilitating the transmission and accumulation of skills or by promoting the development and widespread adoption of new technologies and business practices. Identifying these possible sharing, matching, and learning mechanisms and assessing their relative importance is difficult because they all share the prediction that productivity increases with the scale of an activity at a location.

Just as bigger cities tend to host more productive establishments, they are also increasingly home to a disproportionate share of more educated and particularly productive workers. These sort of differences have grown over time. There is debate about whether big cities, in addition to having more workers at the high end of the skill distribution, also have more workers at the low end, so that the overall distribution of observable skills has greater variance in bigger cities.

Workers in bigger cities not only have higher education on average, they also have higher nominal earnings even after discounting the effect of their greater education. Higher nominal earnings do not necessarily mean that workers are better off in bigger cities, since they also have to pay higher prices for housing. However, it does indicate that they are more productive. If firms in tradable sectors did not find workers in bigger cities more productive, they would locate in smaller cities instead.

All of the above may give the impression that bigger cities hold an increasingly advantageous position relative to smaller cities. And yet, this is not the case. First, because big cities not only feature higher firm productivity and worker earnings, they also have more congestion and substantially higher land and house prices. Furthermore, because as cities have become more different, they have also become more interdependent in many ways. Thus, it would be simplistic to think of urban systems as made of 'winner' and 'loser' cities. Instead, important dynamic patterns have emerged linking cities more closely to each other.

Workers are often willing to pay the higher living cost of bigger cities not just to get higher wages today, but in the hope of learning and acquiring skills that, depending on their own characteristics and luck, they may exploit there or somewhere else. We can find that a substantial part of the productivity benefits of big cities are tied to faster human capital accumulation and learning by tracking workers over time and across cities and measuring how the experience accumulated in different cities increases their earnings. Such an analysis indicates that workers in bigger cities do not have higher initial unobserved ability as reflected in worker fixed-effects. Instead, they obtain an immediate benefit from locating in a big city and also accumulate more valuable experience, which makes their earnings increase faster over time. And those benefits are portable, since the higher value of experience accumulated in bigger cities remains reflected in earnings after workers move away. Overall, about one-half of the gains from working in bigger cities are static and tied to currently working in a bigger city. However, about another half accrues over time as workers accumulate more valuable experience in bigger cities.

Just as there are important life-cycle patterns across cities for workers, these also exist for firms. Many new firms and products start up in cities that are bigger, more educated and more diverse but eventually relocate, making up space for the next generation of new firms and products. In this sense, we can think of diversified cities as 'nursery cities' where learning takes place and specialized cities as the places where the production of mature goods occurs. This can be used to explain why, even if innovation and learning concentrate in a few large and diverse cities, this does not imply that smaller and more specialized cities will disappear. Instead, the diffusion of innovations to exploit them in small specialized cities frees up large and diverse cities to concentrate in continuously feeding the growth process with new ideas.

References

Abdel-Rahman, Hesham M. and Masahisa Fujita. 1990. Product variety, Marshallian externalities, and city sizes. *Journal of Regional Science* 30(2): 165–183.

Agrawal, Ajay, Iain Cockburn, and John McHale. 2006. Gone but not forgotten: Knowledge flows, labor mobility, and enduring social relationships. *Journal of Economic Geography* 6(5): 571–591.

Agrawal, Ajay, Devesh Kapur, and John McHale. 2008. How do spatial and social proximity influence knowledge flows? Evidence from patent data. *Journal of Urban Economics* 64(2): 258–269.

Audretsch, David B. and Maryann P. Feldman. 1996. R&D spillovers and the geography of innovation and production. *American Economic Review* 86(3): 630–640.

Bacolod, Marigee, Bernardo Blum, and William C. Strange. 2009. Skills and the city. *Journal of Urban Economics* 65(2): 127–135.

Bade, Franz-Josef, Eckhardt Bode, and Eleonora Cutrini. 2015. Spatial fragmentation of industries by functions. *Annals of Regional Science* 54(1): 215–250.

Baum-Snow, Nathaniel and Ronni Pavan. 2012. Understanding the city size wage gap. *Review of Economic Studies* 79(1): 88–127.

Berry, Steven and Edward L. Glaeser. 2005. The divergence in human capital levels across cities. *Papers in Regional Science* 84(3): 407–444.

Bettencourt, Luís M. A., José Lobo, Dirk Helbing, Christian Kühnert, and Geoffrey B. West. 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences* 104(17): 7301–7306.

Breschi, Stefano and Francesco Lissoni. 2009. Mobility of inventors and networks of collaboration: An anatomy of localised knowledge flows. *Journal of Economic Geography* 9(4): 439–468.

Brunelle, Cedric. 2013. The growing economic specialization of cities: Disentangling industrial and functional dimensions in the Canadian urban system, 1971–2006. *Growth and Change* 44(3): 443–473.

Bureau of Transport Statistics. 2012. *National Transportation Statistics*. Washington, DC: United States Department of Transportation.

Chandler, Alfred D. 1977. *The Visible Hand: The Managerial Revolution in American Business*. Cambridge, MA: Belknap.

Ciccone, Antonio and Robert E. Hall. 1996. Productivity and the density of economic activity. *American Economic Review* 86(1): 54–70.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2008. Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63(2): 723–742.

Combes, Pierre-Philippe, Gilles Duranton, and Laurent Gobillon. 2012*a*. The costs of agglomeration: Land prices in French cities. Processed, Wharton School, University of Pennsylvania.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, Diego Puga, and Sébastien Roux. 2012*b*. The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica* 80(6): 2543–2594.

Combes, Pierre-Philippe, Gilles Duranton, Laurent Gobillon, and Sébastien Roux. 2010. Estimating agglomeration effects with history, geology, and worker fixed-effects. In Edward L. Glaeser (ed.) *Agglomeration Economics*. Chicago, IL: Chicago University Press, 15–65.

Davis, Donald R. and Jonathan I. Dingel. 2013. The comparative advantage of cities. Processed, Columbia University.

De la Roca, Jorge, Gianmarco I.P. Ottaviano, and Diego Puga. 2014. City of dreams. Processed, CEMFI.

De la Roca, Jorge and Diego Puga. 2017. Learning by working in big cities. *Review of Economic Studies* 84(1): 106–142.

Desmet, Klaus and Esteban Rossi-Hansberg. 2009. Spatial growth and industry age. *Journal of Economic Theory* 144(6): 2477–2502.

Duranton, Gilles and Henry G. Overman. 2005. Testing for localization using micro-geographic data. *Review of Economic Studies* 72(4): 1077–1106.

Duranton, Gilles and Diego Puga. 2001. Nursery cities: Urban diversity, process innovation, and the life cycle of products. *American Economic Review* 91(5): 1454–1477.

Duranton, Gilles and Diego Puga. 2004. Micro-foundations of urban agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2063–2117.

Duranton, Gilles and Diego Puga. 2005. From sectoral to functional urban specialisation. *Journal of Urban Economics* 57(2): 343–370.

Duranton, Gilles and Diego Puga. 2014. The growth of cities. In Philippe Aghion and Steven N. Durlauf (eds.) *Handbook of Economic Growth*, volume 2B. Amsterdam: North-Holland, 781–853.

Eeckhout, Jan, Roberto Pinheiro, and Kurt Schmidheiny. 2014. Spatial sorting. *Journal of Political Economy* 122(3): 554–620.

Ellison, Glenn and Edward L. Glaeser. 1997. Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy* 105(5): 889–927.

Ellison, Glenn, Edward L. Glaeser, and William Kerr. 2010. What causes industry agglomeration? Evidence from coagglomeration patterns. *American Economic Review* 100(3): 1195–1213.

Feldman, Maryann P. and David B. Audretsch. 1999. Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review* 43(2): 409–429.

Fujita, Masahisa and Ryoichi Ishii. 1998. Global location behavior and organizational dynamics of Japanese electronics firms and their impact on regional economies. In Alfred D. Chandler, Jr., Peter Hagström, and Örjan Sölvell (eds.) *The Dynamic Firm: The Role of Technology, Strategy, Organization and Regions*. Oxford: Oxford University Press, 343–383.

Fujita, Masahisa and Hideaki Ogawa. 1982. Multiple equilibria and structural transition of non-monocentric urban configurations. *Regional Science and Urban Economics* 12(2): 161–196.

Glaeser, Edward L., Jed Kolko, and Albert Saiz. 2001. Consumer city. *Journal of Economic Geography* 1(1): 27–50.

Glaeser, Edward L. and David C. Maré. 2001. Cities and skills. *Journal of Labor Economics* 19(2): 316–342.

Greenstone, Michael, Richard Hornbeck, and Enrico Moretti. 2010. Identifying agglomeration spillovers: Evidence from winners and losers of large plant openings. *Journal of Political Economy* 118(3): 536–598.

Henderson, J. Vernon. 2003. Marshall's scale economies. *Journal of Urban Economics* 53(1): 1–28.

Henderson, J. Vernon and Yukako Ono. 2008. Where do manufacturing firms locate their headquarters? *Journal of Urban Economics* 63(2): 431–450.

Holl, Adelheid. 2004. Transport infrastructure, agglomeration economies, and firm birth: Empirical evidence from Portugal. *Journal of Regional Science* 44(4): 693–712.

Holmes, Thomas J. 1999. Localization of industry and vertical disintegration. *Review of Economics and Statistics* 81(2): 314–325.

Holmes, Thomas J. and John J. Stevens. 2004. Spatial distribution of economic activities in North America. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2797–2843.

Hummels, David. 2007. Transportation costs and international trade in the second era of globalization. *Journal of Economic Perspectives* 21(3): 131–154.

Jacobs, Jane. 1969. *The Economy of Cities*. New York: Random House.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics* 108(3): 577–598.

Kelley, Maryellen R. and Susan Helper. 1999. Firm size and capabilities, regional agglomeration, and the adoption of new technology. *Economics of Innovation and New Technology* 8(1–2): 79–103.

Kim, Sukkoo. 1999. The rise of multiunit firms in US manufacturing. *Explorations in Economic History* 36(4): 360–386.

Lainé, Frédéric and Carole Rieu. 1999. La diversité industrielle des territoires. Premiere 650, Institut National de la Statistique et des Etudes Economiques.

Melitz, Marc and Gianmarco I. P. Ottaviano. 2008. Market size, trade and productivity. *Review of Economic Studies* 75(1): 295–316.

Michaels, Guy and Ferdinand Rauch. 2015. Resetting the urban network: 117–2012. Processed, London School of Economics.

Michaels, Guy, Ferdinand Rauch, and Stephen J. Redding. 2012. Urbanization and structural transformation. *Quarterly Journal of Economics* 127(2): 535–586.

Moretti, Enrico. 2011. Local labor markets. In Orley Ashenfelter and David Card (eds.) *Handbook of Labor Economics*, volume 4. Amsterdam: Elsevier, 1237–1313.

Moretti, Enrico. 2012. *The New Geography of Jobs*. New York, NY: Houghton Mifflin Harcourt.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura. 2014. Localized knowledge spillovers and patent citations: A distance-based approach. *Review of Economics and Statistics* 96(5): 967–985.

No, Joung Yeo Angela. 2003. Agglomeration effects in the diffusion of advanced manufacturing technologies. Processed, University of Toronto.

Overman, Henry G. and Diego Puga. 2010. Labour pooling as a source of agglomeration: An empirical investigation. In Edward L. Glaeser (ed.) *Agglomeration Economics*. Chicago, IL: Chicago University Press, 133–150.

Rappaport, Jordan. 2007. Moving to nice weather. *Regional Science and Urban Economics* 37(3): 375–398.

Rodríguez-Pose, Andrés and Rune Dahl Fitjar. 2013. Buzz, archipelago economies and the future of intermediate and peripheral areas in a spiky world. *European Planning Studies* 21(3): 355–372.

Rosenberg, Nathan and Manuel Trajtenberg. 2004. A General Purpose Technology at work: The Corliss steam engine in the late 19th century US. *Journal of Economic History* 64(1): 61–99.

Rosenthal, Stuart S. and William Strange. 2004. Evidence on the nature and sources of agglomeration economies. In Vernon Henderson and Jacques-François Thisse (eds.) *Handbook of Regional and Urban Economics*, volume 4. Amsterdam: North-Holland, 2119–2171.

Rosenthal, Stuart S. and William C. Strange. 2001. The determinants of agglomeration. *Journal of Urban Economics* 50(2): 191–229.

Sonn, Jung Won and Michael Storper. 2008. The increasing importance of geographical proximity in knowledge production: An analysis of us patent citations, 1975–1997. *Environment and Planning A* 40(5): 1020–1039.

Stigler, George J. 1951. The division of labor is limited by the extent of the market. *Journal of Political Economy* 59(3): 185–193.

Sveikauskas, Leo. 1975. Productivity of cities. *Quarterly Journal of Economics* 89(3): 393–413.

Thompson, Peter and Melanie Fox-Kean. 2005. Patent citations and the geography of knowledge spillovers: A reassessment. *American Economic Review* 95(1): 450–460.

World Bank. 2009. *World Development Report: Reshaping Economic Geography*. Washington, DC: World Bank.

Yates, JoAnne. 1989. *Control Through Communication: The Rise of System in American Management*. Baltimore, MD: Johns Hopkins University Press.

Yates, JoAnne. 1991. Investing in information: Supply and demand forces in the use of information in american firms, 1850–1929. In Peter Temin (ed.) *Inside the Business Enterprise: Historical perspectives on the use of information*. Chicago, IL: University of Chicago Press, 117–154.