Agglomeration and Jobs

33

Gilles Duranton

Contents

Abstract

This chapter discusses the literature on agglomeration economies from the perspective of jobs and job dynamics. It provides a partial review of the empirical evidence on agglomeration externalities; the functionality of cities; the dynamic relationship between cities, jobs, and firms; and the linkages between cities. We provide the following conclusions. First, agglomeration effects are quantitatively important and pervasive. Second, the productive advantage of large cities is constantly eroded and needs to be sustained by new job creations and innovations. Third, this process of creative destruction in cities, which is fundamental for aggregate growth, is determined by the characteristics of urban systems and broader institutional features. We highlight important differences between developing countries and more advanced economies. A major challenge for developing countries is the transformation of their urban systems into drivers of economic growth.

G. Duranton
Department of Economics, University of Toronto, Toronto, ON, Canada
e-mail: duranton@wharton.upenn.edu

M.M. Fischer, P. Nijkamp (eds.), *Handbook of Regional Science*,
DOI 10.1007/978-3-642-23430-9_33,

33.1 Introduction

This chapter reviews the literature on agglomeration economies from the perspective of jobs and labor markets.

In cities, jobs are more productive because of agglomeration effects. These take place through a variety of channels: resource sharing, quicker and better matching, and greater knowledge spillovers. Section 33.2 provides a discussion of these issues. The bottom line is straightforward; cities have a positive effect on productivity and wages.

More productive urban jobs however do not come in a void. Section 33.3 broadens the discussion to job creation and firm dynamics in cities. More productive jobs in cities need to be created. Innovation, entrepreneurial activity, and firm growth all play a crucial role in this respect. Adding to this, more productive jobs do not remain more productive forever. This productivity advantage is constantly eroded and needs to be constantly re-created. The creative destruction process, that is, more firm entry and exit and higher portion of innovative young firms, is also fundamental.

In turn, the dynamics of firms and jobs in cities is shaped by the broader characteristics of urban systems. In Sect. 33.4, we highlight major differences between cities in developing countries and more advanced economies. In short, the urban system of many developing countries acts as a brake on economic growth. A major challenge for the countries is the transformation of their urban systems into drivers of economic growth. More specifically, cities in developing countries appear to be far less functionally specialized than cities in more advanced economies. This hampers the dynamism of the largest cities in developing countries which are burdened by many ancillary activities. These activities add to urban crowding without adding to agglomeration benefits. Better infrastructure, in particular better transportation infrastructure, and a reduction in favoritism toward large cities may be a way to remedy these problems. Policies to foster job creations directly may be tempting, but their record in more advanced economies is unsatisfactory. In addition, developing cities also function less efficiently and face challenges that differ from those of cities in more advanced economies. An appropriate management of the transition to full urbanization, a strengthening of urban governance, a reduction in labor market duality, and a reduction or the full elimination of land market duality are key challenges that must be tackled for developing cities to take the full advantage of agglomeration effects and foster aggregate growth.

33.2 Cities, Worker Productivity, and Wages

Cities enjoy a productive advantage over rural areas, and this advantage is larger for larger cities. The positive association between various measures of productivity and urban scale has been repeatedly documented. That larger cities obtain higher scores on many productivity metrics from wages to output per worker, or the total factor

productivity of firms is now beyond doubt. Most of the studies reviewed by Puga (2010) find an elasticity of wages or firm productivity with respect to city employment or urban density between 0.02 and 0.10. As shown by Henderson (2005), these findings also hold widely in cities in developing countries.

More formally, this type of work involves regressing an outcome variable by location on a measure of agglomeration. In the early literature, the typical regression of choice involved using output per worker as dependent variable and city population as explanatory variable. In the early 1990s, authors often employed more indirect strategies and started to use variables such as employment growth or firm creation as outcome measures. More recently, the literature has moved to microdata and returned to more direct outcome measures, namely, the total factor productivity of firms and wages. More precisely, recent studies estimate a regression like

$$\log w_{ic(i)} = \alpha \log Pop_{c(i)} + \eta_{c(i)} + u_i + \varepsilon_{ic(i)}, \tag{33.1}$$

where c denotes cities and i denotes individuals or groups of individuals. The dependent variable is w the wage, and the explanatory variables are $\log Pop$ the log of population as a measure of urban scale, η a city effect (usually proxied through a number of control variables at the city level), and u an individual effect (often proxied through observable individual characteristics). Finally, ε is an error term. The estimated value of the coefficient of interest, α, is usually positive and significant. Similar regressions can be proposed for firm data using measures of firm level productivity and firm characteristics.

After Ciccone and Hall (1996), density has often been favored relative to population since it appears to yield more reliable results. The reason is probably that density-based measures of agglomeration are more robust to zoning idiosyncrasies. For instance treating Washington and Baltimore as one big consolidated metropolitan area or two separate cities makes a big difference to their employment count but only little difference to density.

After asserting this robust statistical association between productivity outcomes and agglomeration, the first question regards whether the estimated coefficient α in the regression described by Eq. (33.1) reflects the causal effect of agglomeration on wages. An examination of Eq. (33.1) reveals three possible sources of bias. They all come from the fact that, as highlighted by the notations in Eq. (33.1) above, the measure of agglomeration is indexed by $c(i)$, that is, the city c is *chosen* by worker i. Ideally, one would like to compare the same workers across the cities that they have chosen and those that they have not chosen. In absence of randomized experiments, this is not possible. Greenstone et al.'s (2010) quasi experiment on "million dollar plants" is what comes closest to this ideal for firms' location choices.

The first source bias is the possible link between city effects (which are not observed directly) and the variable of interest, city population, or density. Put differently, the "quantity of labor" may be endogenous, and it is reasonable to expect workers to go to more productive cities. A possible solution to this problem

is to use instruments for city population or density as Ciccone and Hall (1996). These instruments need to predict current population patterns but must be otherwise uncorrelated with city productivity. Deep historical lags such as population from 200 years ago or soil characteristics can do the job. Studies using this type of approach typically find that correcting for the endogeneity of population has only a mild downward effect on the estimation of the coefficient of interest α.

The second main identification problem in the estimation of Eq. (33.1) regards a possible correlation between the measure of city population and individual effects. That is, the quality of labor may be endogenous, and we expect more productive workers to reside in larger cities. A first possible solution to this problem is to control for an extensive set of individual characteristics. A more drastic solution is to use (whenever possible) the longitudinal dimension of the data and impose worker fixed effects as Combes et al. (2008). The endogenous quality of labor seems to be an important source of bias in the estimation of Eq. (33.1). The estimated value of α is typically reduced by 30–50 % using extensive individual controls or worker effects. This said, one needs to be careful. Imposing worker effects improves the quality of the estimation, but it is not a perfect solution since it assumes that mobility is exogenous.

Related to this last issue, the third source of bias in the estimation of Eq. (33.1) is the possibility of a correlation between the error term and the measure of city population of interest. If, for instance, workers move more easily from large cities to small cities than the opposite in case of a good external wage offer, this will create another source of bias which in this particular situation leads to an underestimate of agglomeration economies. No satisfactory solution to this problem has been proposed so far.

At this point, the conclusion of the agglomeration literature is that there is a causal static effect of cities and urbanization on wages in more advanced economies but that this effect represents only about half the measured association between city population or density and wages (or alternative measures of productivity). The rest of the association between population or density and wages reflects the sorting of more productive workers in larger and denser cities and, to a lesser extent, reverse causality and workers moving to more productive places. Recent investigations that tackle the concerns mentioned above find agglomeration elasticities around 2 %. They thus suggest rather modest static effects of cities on productivity. The literature from developing countries often uses less sophisticated approaches but finds results that are comparable and, if anything, indicative of moderately stronger agglomeration effects.

After questioning its causal aspect, the second key question about the estimation of agglomeration effects regards their sources. When asking about the "sources" of agglomeration, the literature frequently confuses two separate questions. The first is about which markets are affected by these agglomeration effects, and the second is about which mechanisms actually occur. Regarding the "where" question, it is customary to distinguish the markets for (intermediate) goods, the market for labor, and the (absent) market for ideas and knowledge. In terms of mechanisms, we often distinguish between sharing, matching, and learning mechanisms.

"Sharing" is about the many possible benefits from the mutualization of specialized input providers, the diversity of local goods, the division of labor, or the risks. "Matching" is about the greater probability of finding another party such as a worker, an employer, a supplier, or an investor and the greater quality of the match with that party. Finally, "learning" is about the better generation, diffusion, and accumulation of knowledge. The latter set of mechanisms is regularly referred to as knowledge spillovers.

Because of the wide variety of possible mechanisms and the markets where they can take place, the literature that investigates the sources of agglomeration benefits is much more heterogeneous than the literature that attempts to measure the overall benefits from agglomeration. The latter naturally coalesces around the estimation of Eq. (33.1).

First, there is a diversity of work which provides evidence of an association between some aspect of agglomeration such as a particular mechanism or market and measures of agglomeration such as city size. Let us take only a few recent examples (see Puga 2010, for a more exhaustive discussion). Taken together, these studies are suggestive that many of the agglomeration mechanisms described by the theoretical literature are at work in a variety of markets.

This conclusion must be taken cautiously, however. Establishing the direction of causality in this type of work is even harder here than when attempting to measure the overall effects of agglomeration. To understand this point and the pitfalls associated with this type of work, let us use the analysis of Charlot and Duranton (2004) on workplace communication. They show that communication is associated positively with city size and with wages. This leads them to conclude that communication spillovers could account for up to a quarter of agglomeration benefits. However, this finding could be explained in part by the greater sorting of good communicators in larger cities. This is the equivalent of the quality-of-labor bias discussed above. This worry can be reduced by comparing movers and stayers in cities as Charlot and Duranton (2004) do. It is difficult to eliminate it entirely though. In addition, one also needs to show that greater communication in cities is not the by-product of another agglomeration force. Workers in larger cities may communicate more because firms outsource more of their output. This requires some coordination. In such a case, the real source of agglomeration benefits may be input–output linkages, not communication spillovers. To go round this problem, Charlot and Duranton (2004), who use rich firm level data, suggest instrumenting workplace communication by measures of organizational changes such as a flattening of the hierarchy. These changes typically increase the need for horizontal communication. This type of instrument is nonetheless valid only if changes in organization are unrelated to other sources of agglomeration benefits such as labor pooling or input–output linkages. That firm reorganization affects worker communication behavior but has no direct effect on recruiting practices, or outsourcing is plausible but not certain. More generally, studies that focus on one particular source of agglomeration face a major missing variable problem: The other sources of agglomeration are absent from the regression even though they are expected to be correlated with both wages (or other productivity measures) and measures of agglomeration such as city size.

Given how difficult it is to measure many aspects of agglomeration and given also that the list of possible agglomeration sources is open, considering all sources of agglomeration in one regression is not a feasible option. A more reasonable path forward is, following Ellison et al. (2010), to consider several classes of agglomeration sources in the same approach. Ellison et al. (2010) assess how much labor pooling, input–output linkages, and spillovers account for co-agglomeration between industries in the USA. They use a measure of industry co-agglomeration and find more co-agglomeration among (i) industries that buy from each other, (ii) industries that use a similar workforce, and (iii) industries that share a common scientific base. To reduce the possibility that co-agglomerated industries end up buying from each other or using similar workers because of their proximity, they instrument their US measures of input–output linkages and labor pooling using corresponding UK data. Of course, if the biases are the same in the UK as in the USA, these instruments are of limited value. Another caveat is that input–output linkages are possibly more easily measured using input–output matrices than spillovers using patent citations. This can also lead to biased estimates since a positive correlation with both linkages and spillovers is likely to be picked up mainly by the better-measured linkage variable. This said, Ellison et al. (2010) confirm that the three motives for agglomeration they consider are at play with input–output linkages playing a more important role.

Even if we abstract from the uncertainty around those results, the notion that several mechanisms, each operating in several markets, contribute to agglomeration benefits is problematic for policy. At their heart, agglomeration benefits rely on market failures associated with the existence of small indivisibilities with sharing mechanisms, thick market effects with matching mechanisms, and uncompensated knowledge transfers with learning mechanisms. That is, there are possibly many market failures at play in many markets. In turn, this implies that there may be no hope of fostering agglomeration economies through a small number of simple policy prescriptions.

Before broadening the discussion, there are four further features of agglomeration that have implications for workers and jobs in cities.

The first is the issue of the sectoral scope of agglomeration and whether agglomeration effects accrue mostly within or across sectors. Agglomeration effects within sectors are referred to as localization economies and between sectors as urbanization economies. When estimating a more general version of Eq. (33.1) that accounts for both city size or density and the degree of same sector specialization, extant research has found evidence of both localization and urbanization effects. There are two interesting nuances. The first is the presence of significant heterogeneity across industries. This heterogeneity follows an interesting pattern as it appears that more technologically advanced industries benefit more from urbanization economies whereas more mature industries benefit more from localization economies. Second, the calculations of Combes et al. (2008) indicate that in France the benefits from localization economies are smaller than those of urbanization economies and mostly uncorrelated with local wages. Put differently, increased local specialization has only small benefits and does not contribute to making workers richer.

The second extra feature of agglomeration is the notion that not all workers benefit equally from urban scale. Equation (33.1) estimates an "average" agglomeration effect. As highlighted by, among others, Glaeser and Resseger (2010), agglomeration effects appear stronger for more educated workers in the USA. Higher returns in larger cities should in turn provide stronger incentives to more skilled workers to locate there. Hence, these results are consistent with the well-documented fact that workers in larger cities in more advanced economies tend to be more educated and better skilled (e.g., Combes et al. 2008).

Next, while not all workers benefit equally to agglomeration effects, it also appears that not all workers contribute equally to these effects either. There is a large literature on human capital externalities suggesting that workers enjoy higher wages when surrounded by more educated workers. Estimates of external returns to education are typically between 50 % and 100 % the corresponding estimates of private returns to education, in particular for university graduates. These findings are robust to a number of estimation concerns and suggestive of large effects. It is beyond the scope of this chapter to review this literature extensively. See instead Moretti (2004) for an in-depth survey.

Finally, there is also emerging evidence from US and European data that wage growth also depends on city size/density. To show this, one can estimate a regression along the lines of Eq. (33.1) but use wages in first difference instead of in levels as dependent variable:

$$\Delta_{t+1,t} \log w_{ic(i)} = \alpha \log Pop_{c(i)t} + \eta_{c(i)} + u_i + \varepsilon_{ic(i)t} \tag{33.2}$$

where Δ is used to note time differences between t and $t + 1$. Among a number of papers, De la Roca and Puga (2012) confirm that wage growth is stronger in larger cities.

Because the structure of Eq. (33.2) is the same as that of Eq. (33.1) for the static estimation of agglomeration economies, it suffers from the same drawbacks. First, the association between wage growth and agglomeration could be explained by the sorting of workers with faster wage growth in larger cities. This could occur because "fast learner" tends to locate in larger cities or because the wage of workers who are predominantly located in larger cities (such as more educated workers) tends to increase faster. Following the same sort of fixed effect strategy described above and applying that to a regression like Eq. (33.2), Freedman (2008) nonetheless shows that this type of result holds even after controlling for the fact that some workers may experience higher wage growth independently of their location.

Although the result that wages grow faster in cities is frequently interpreted as evidence about faster learning in cities and knowledge spillovers, the mechanisms that drive it are unclear. Just like regressing wages in levels on a measure of urban scale in Eq. (33.1) does not tell us anything about the sources of static agglomeration economies, regressing wage growth on urban scale in Eq. (33.2) is equally uninformative about the sources of agglomeration dynamics. Interestingly, Wheeler (2008) shows that young workers tend to change job more often in larger cities, while the opposite holds for old workers. This type of evolution is consistent with

a matching model where workers can find their "ideal match" faster in larger cities and then stick to it. Such mechanism could explain both faster wage growth and eventually higher wages in larger cities.

Evidence about learning in cities can come from the fact that workers retain some benefits from agglomeration after they leave their city. De la Roca and Puga (2012) confirm this on Spanish data. Their findings suggest the existence of both a level effect of cities on wages (of the same magnitude as those discussed above) and a dynamic effect. Over the long run, workers in large cities seem to gain about as much from both effects.

To sum up, this discussion of agglomeration economies which focuses mainly on workers and jobs reaches a number of interesting conclusions. First, larger cities make workers more productive. There is both a static and a dynamic component to these gains. A static elasticity of wages with respect to city population of 0.03 implies that a worker receives a 23 % higher wage when moving from a tiny city with population 5,000 to a large metropolis with a population of five million. Over time, dynamic effects could make this urban premium twice as large. While long-run gains close to 50 % are not miraculous, they are nonetheless sizeable.

In terms of policy implications, the temptation to "foster agglomeration effects" should be resisted. We are too far from knowing enough about the sources of agglomeration to implement any meaningful policy, not to mention the great heterogeneity in who gains from and who contributes to agglomeration gains. It remains nonetheless that the economic gains from urbanization are significant and urbanization should be embraced rather than resisted.

33.3 Firm Dynamics Within Cities

This higher productivity of jobs in cities is only one facet of the issue. Jobs are usually viewed as a veil when we model production in theoretical models. In practice, higher labor productivity is associated with doing different things and doing them differently. That is, to receive higher wages, workers need "better jobs." Firm dynamics is often the vector of these changes. More specifically, let us examine several aspects of firm dynamics in cities: innovation, firm creation and growth, and factor allocation and reallocation across firms.

Starting with innovation, the first salient feature of the geography of innovative activity is that research and innovation is much more concentrated than production in most industries. Interestingly, this tendency seems particularly strong for industries that are more intensive in skilled labor and in research and development. It is also the case that this concentration of research and development typically takes place in large metropolitan areas.

These location patterns for innovative activity are consistent with the notion that cities have a positive effect on innovation just like they have on wages. More direct evidence can be found in Feldman and Audretsch (1999) and Carlino et al. (2007). To measure innovation, Feldman and Audretsch (1999) make a count of all new product innovations in US metropolitan areas for a broad set of technologies and

sectors in 1982. They find no evidence of urban scale effects but find that same sector specialization is strongly negatively associated with innovation whereas a diversity of employment in technologically related industries is strongly positively associated with innovation. They also find strong positive innovation effects associated with the presence of smaller establishments.

Using the number of patents per capita as dependent variable, Carlino et al. (2007) find evidence of strong agglomeration effects for innovative activity. Their estimate of the elasticity of patenting per capita with respect to employment density is 0.2. This is several times the estimates reported above for the corresponding elasticity of wages. Interestingly, Carlino et al. (2007) also find that this elasticity of innovation with respect to employment density or population size is not constant across the urban hierarchy. Patenting per capita appears to peak at around 5,700 jobs per square kilometer or a city population size slightly below a million.

While this evidence is highly suggestive that cities affect innovation, there is, to the best of our knowledge, no work which focuses on the effects of innovative activity in cities such as its effects on urban growth. Regressing urban population growth on innovative activity would raise some obvious identification concerns. In addition, simple theoretical argument suggests that the effect of innovation on urban growth need not be positive. Obviously, product innovation in the form of either an entirely new product or the capture of an established product from another location is expected to add to a city's employment. However, process innovation within a city can cut both ways. Employment will increase with process innovation only if greater productive efficiency and lower prices lead to a more than proportional increase in demand. In the opposite case, process innovation will imply a contraction of local employment. Remarkably, Carlino et al. (2007) show that Rochester, Buffalo, Cleveland, St Louis, and Detroit are all highly innovative cities. This suggests that, to some extent, the demise of these cities may be attributed to the fact that labor productivity increased much faster than demand in their industries.

Finally, innovative activity appears to change the nature of jobs in the cities where it takes place. As shown by Lin (2011), cities that patent more tend to have a greater proportion of what he labels "new work," that is, jobs that did not exist a few years before. New work is also fostered by a greater proportion of educated workers and a diversity of industries, two other attributes of large cities.

To conclude on the links between innovation and cities, extant literature supports the notion that cities affect innovation either because of their sheer population size or because of the (diverse) structure of their production activities. The evidence about the effect of innovation on cities is more complex. Innovation *within a given city* affects the proportion of workers in new work. Other effects are either ambiguous or poorly documented. As we show below, further insights about the effects of innovation on cities can be gained by looking across cities.

Entrepreneurship is also closely associated with cities in several ways. First, cities affect entrepreneurship just like they affect wages and innovation. In a comprehensive analysis of the determinants of employment in new manufacturing start-ups across sectors in US cities, Glaeser and Kerr (2009) generate a rich harvest of facts. The first is the existence of scale economies. As a city grows larger,

employment in new start-ups in this city increases more than proportionately. Depending on their specification, Glaeser and Kerr (2009) find an elasticity of employment in new start-ups per capita with respect to city scale between 0.07 and 0.22. City population, city-industry employment, and sector effects explain around 80 % of the variation in start-up employment across cities and sectors.

Glaeser and Kerr (2009) also find that the presence of many small suppliers has a strong effect on employment in start-ups. In addition, they also find evidence of mild Marshallian effects associated with input–output linkages, labor market pooling, and spillovers. Finally, city demographics only has a limited explanatory power just like their measure of "entrepreneurial culture."

The other key feature about the supply of entrepreneurs is that there is a strong local bias in entrepreneurship. Entrepreneurs tend to create their start-up in the place where they were born and/or where they have lived and worked before becoming entrepreneurs. This important fact has been documented by Figueiredo et al. (2002) for Portugal and Michelacci and Silva (2007) for Italy and the USA. This finding has been confirmed by several other studies in developed economies. Figueiredo et al. (2002) also show that when entrepreneurs chose a new location, this choice is strongly governed by agglomeration economies and a proximity to large cities.

After looking at the urban determinants of entrepreneurship, we now turn to the effects of entrepreneurship on their cities. It has been shown repeatedly that entrepreneurship plays a key role in urban evolutions. The key fact here is that growth in a city and sector over a period of time is strongly correlated with the presence of small establishments in that city and sector at the beginning of the period. This fact was first documented by Glaeser et al. (1992) and has been confirmed for other countries and time periods by many other studies.

Just like with many of the correlations discussed above, the strong link between small firms and employment growth raises a key identification concern about the direction of causality. However, this issue has been neglected by the literature. This is perhaps because the standard regression uses growth over a period as dependent variable and establishment size *at the beginning of the period* as explanatory variable. However, using a predetermined variable as explanatory variable in a regression does not guarantee its exogeneity. Local entrepreneurs could enter in large numbers in a city and sector if they foresee strong future demand. That expectations of future growth should trigger entry today is only natural. That is the nature of business.

To resolve this identification problem, it is difficult to think of instruments that would predict establishment size in a city and sector but be otherwise uncorrelated with subsequent growth. To clarify the meaning of the relationship between small establishments and high subsequent growth, Glaeser et al. (2010) do something quite different. They look at whether the presence of many small firms in a city and sector is driven by the demand for entrepreneurship or its supply. To the extent that the demand for entrepreneurship can be captured by higher sales per worker, this does not appear to be the case. They also find limited evidence about the importance of lower labor costs or entrepreneurs sorting into high amenity cities. They find

stronger evidence about the importance of the proportion of university graduates (particularly in more skilled industries), but that still does not explain away the effect of having lots of small establishments. While still preliminary, this type of evidence points at some unspecified supply effects. More entrepreneurial cities happen to have a greater supply of entrepreneurs, and the literature has thus far been unable to trace this further.

Turning finally to factor allocation and reallocation, the literature that examines these issues makes two important claims. The first is that a large fraction of productivity growth at the country level can be accounted for by the reallocation of factors from less productive to more productive firms. A large share of productivity growth can be accounted for by a churning process where low-productivity firms are replaced by new and more productive start-ups. These important findings have been confirmed for many countries (Bartelsman et al. 2004).

The second important claim made by the reallocation literature is that "misallocation" can account for a large share of existing productivity differences across countries. To understand this point better, consider the influential work of Hsieh and Klenow (2009). They first note that, in equilibrium, the marginal product should be equalized across firms. If the demand for the varieties produced by firms has a constant elasticity of substitution, this implies an equalization of the product of their price by their "true productivity" (which is the ability of firms to produce output from inputs). This – price times true productivity – product is what is estimated as "total factor productivity" in most productivity exercises. We may call this second quantity "apparent productivity." Obviously, the firms' apparent productivities are never equalized in real data. Hsieh and Klenow (2009) interpret this as evidence of factor misallocation. Taking the highly dispersed distribution of manufacturing productivity in China and India, they calculate very large potential costs from such misallocation. Acknowledging that a perfectly efficient allocation may be impossible, they compute that the productivity gains for manufacturing in China and India would still be of about 50 % if their level of misallocation could be reduced to that observed in the USA.

To the best of our knowledge, there is no study that would attempt to relate greater churning/reallocation at the firm level and higher productivity growth at the urban level. However, there is a strong suspicion that larger cities should exhibit more churning. This is because, as already argued, larger cities are more innovative, experience more entry and exit, and have a greater fraction of their workforce in "new work." At the same time, there is no indication that this greater amount of churning in larger cities is associated with higher productivity growth in those cities unlike what occurs at the country level.

We actually know little about productivity growth in cities. According to Lin (2011), the greater proportion of workers employed in new work in larger cities is not associated with faster productivity growth. In a rare study of the broader determinants of productivity growth in Italian cities, Cingano and Schivardi (2004) highlight the importance of both specialization and employment size. But given that specialization and employment size are negatively correlated, their positive effects arguably cancel out. Hence, more churning does not appear to lead to faster productivity growth in cities.

To confirm this conclusion, note that workers are somewhat mobile across cities. Then more churning associated with faster productivity growth in larger cities should imply a divergence in population growth rates. There is no evidence of such divergence. This lack of result regarding the link between churning and productivity should not be taken as negative evidence against the reallocation literature. As argued in the next section, it is possible that reallocation does not take place within cities but also across cities.

Turning to the second claim about misallocation, Combes et al. (2011) show that the distribution of firm productivity is unambiguously more dispersed in larger cities in France. In the framework of Hsieh and Klenow (2009), that would be interpreted as greater misallocation in larger cities. This seems hard to believe. The evidence about static agglomeration effects discussed above is instead best interpreted as agglomeration economies leading to a better allocation of resources (in a broad sense) in larger cities. When performing a productivity decomposition, Combes et al. (2011) find a similar covariance between establishment size and productivity in large and small cities which suggest a similar level of efficiency in the allocation of factors to firms across cities of all sizes.

To sum up, the evidence about firm dynamics and cities presented in this section is puzzling. Larger cities seem to be more innovative, be more entrepreneurial, experience more churning and reallocation, and generally enjoy a greater "economic dynamism." At the same time, they do not appear to enjoy most of the benefits associated with such dynamism since neither productivity nor population appears to increase faster in larger cities. Of course, these conclusions need to be taken cautiously given the paucity of study, including their complete absence for cities in developing countries.

33.4 City Functionality, Urban Systems, and Policies

The answer to the apparent puzzle raised above is that when thinking about economic growth, it is wrong to think of cities as self-contained units. Cities are best viewed as small open economies which interact a lot with other cities and rural areas. They are part of an "urban system." This implies that innovation, churning, and reallocation are best studied across the entire system of cities.

Starting with innovation, recall that larger cities offer many advantages for both product and process innovation. More specifically, as highlighted by many, cities favor the circulation and cross-fertilization of ideas. This naturally leads to more product innovations, and this is consistent with the evidence of Feldman and Audretsch (1999) discussed above. For process innovation, Duranton and Puga (2001) underscore the greater availability of intermediate goods in large cities which allows firms to proceed through trial and error at a faster pace. Put differently, the greater ability of larger cities to innovate may just be another manifestation of agglomeration economies. The key difference with many static aspects of agglomeration economies such as thicker local labor markets is that, with dynamic effects, co-location is not needed all the time. More precisely, spillovers may matter to develop an innovation, but after this is done, co-location is no longer needed.

Quite the opposite, larger cities are more expensive places to produce. After the dynamic benefits from agglomeration have been exploited, it can make sense for firms to relocate. Often, the entire firm does not need to relocate since it is only the production of particular products that is concerned.

Patterns of establishment relocations in France are highly consistent with this type of product cycle. As shown by Duranton and Puga (2001), about 75 % of French establishments that relocate do so from a city with above-median diversity to a city with below-median diversity and above-median specialization in the same sector. In addition, as documented by Fujita and Ishii (1998), large Japanese multinationals in the electronic sector produce their newest products in "trial" plants near Tokyo and Osaka. Less recent products are produced in rural locations in Japan while even older generations of their products are manufactured in less advanced countries in Asia. Hence, as their products mature, firms still search for agglomeration economies but will put a greater weight on the benefits of specialization. Large cities act as nurseries for new goods and new products. Once mature, new goods and products are best produced in more specialized places.

Cities are also specialized by sector. However, this tendency, while still present in the data, has diminished over time as documented by Duranton and Puga (2005). The same authors also document a rise in the functional specialization of cities with the emergence of cities specialized into management-type functions, whereas others specialize more into production activities. This rise in functional specialization is rationalized by Duranton and Puga (2005) in a model where lower communication costs make it easier for firms to separate management from production. Since these activities benefit from very different types of agglomeration economies, such separation is beneficial, provided the cost of separating activities is low enough. In turn, this separation of activities reinforces the functional specialization of cities.

These multiple dimensions of specialization are part of well-functioning urban systems in more advanced countries. Adding to this, the notion of cities being specialized by functions and activities is not static. The process of continuous location and relocation of economic activity is a crucial aspect of the growth of those activities. To take a simple example, when George Eastman developed a new revolutionary technology in the photographic industry in Rochester, the latter relocated from New York to Rochester. Then, much later, as the technology developed by Eastman got itself superseded by the digital revolution, Rochester lost its status of capital of the photographic industry.

That different cities specialize into different functions and are able to change their specialization after negative shocks presupposes a fair amount of "mobility" across cities. The first important dimension of mobility regards goods and services. It would make little sense for cities to narrowly specialize in an activity if its output cannot be exported. Continuously changing patterns of specialization also require labor mobility. For instance, Kerr (2010) documents that after "breakthrough" innovations, more innovations tend to take place in the same location for the same technology. This growth in patenting, in turn, depends on the mobility of scientists and engineers. Interestingly, the adjustment appears faster for technologies that depend more heavily on immigrant inventors who are more mobile.

While the foregoing discussion describes well what happens within the urban system of more advanced economies, it is a far less resemblant depiction of the situation of cities in developing countries. For instance, most very large cities in developing countries are still major manufacturing centers, whereas manufacturing production is mostly absent from the largest cities of Europe and North America. This lack of urban differentiation may be at the root of the problem. Urban systems in developing countries may be much less efficient than in more advanced countries because cities are much less differentiated in terms of functions.

More specifically, this lack of differentiation in urban functionality may hamper the dynamism of cities in developing countries. The largest cities there are burdened by many ancillary activities such as basic manufacturing and call centers. These add to urban crowding without adding to agglomeration benefits. On the other hand, smaller cities in developing countries often lag far behind, and getting some of these ancillary activities would be crucial for their development.

This said, a lack of well-functioning urban systems – however important (and neglected in urban policy) – is not the only cause for the lower efficiency of cities in developing countries relative to their counterparts in advanced economies. Nonurban factors such as weak national institutions and poor technology certainly play a role. Urban factors which hinder the functional differentiation of cities also have a direct negative effect on the efficiency of cities. For instance, as we discuss below, high transportation costs limit the specialization of cities by reducing their ability to trade. At the same time, even if we abstract from these effects, high transportation costs also affect the price of goods purchased by local consumers and reduce market access for local producers.

In the rest of this section, we examine a number of urban factors that both reduce the efficiency of the urban system as well as the efficiency of cities directly. Cities in developing countries are often acting as a brake on growth, whereas they should be a key driver of economic development.

The first key difference between cities in developing and more advanced countries regards the functioning of their labor market. In most developing countries, there is a well-known duality in the labor market which usually comprises a large informal sector alongside the formal sector. Aside from its detrimental implications for workers in the informal sector, this duality hinders urban development in several ways. First, it has been accused of inducing too much migration toward the largest cities where most of the formal sector is located. Duality may also limit mobility across cities since jobs in the informal sector tend to be filled by word-of-mouth through social connections which are missing to newcomers. High barriers to "good" jobs in the formal sector may also hold back the incentives of workers to improve their skills locally and thus limit the scope of agglomeration benefits.

To mitigate the effects of labor market duality, three broad types of policies can be envisioned. The first is to improve the working of labor markets. While this objective is certainly laudable, a discussion of this class of policies would certainly go beyond the scope of this chapter.

The second type of policy is to foster local job creation through "place-based" policies. Such policies typically involve tax exemptions or subsidies associated

with job creation within well-defined (and often tightly circumscribed) areas. These tools are frequently used to try to reduce the unemployment rate of the residents of poor areas in more advanced economies. While the labor market failures in developed and developing countries differ and the scale at which such policies might be implemented in developing countries may be much broader than poor neighborhoods of "rich" cities, there may be useful lessons to learn from the recent North American and European literature evaluating those policies. Simply put, the general record of place-based policies is in doubt. Detailed evaluations of particular policies are usually negative (Glaeser and Gottlieb 2008).

The third class of policies attempts to foster job creations in a particular locality by helping firms in a given sector. These policies are usually referred to as "cluster" policies and follow from the work of Michael Porter (1990). They often entail the development of subsidized supportive institutions and infrastructure using public subsidies and various types of fiscal incentives. The review of the literature in Puga (2010) implies negative conclusions about the possible benefits of cluster policies.

The second key difference between cities in developing and more advanced countries regards the functioning of their land market. Like labor markets, land markets in developing cities are characterized by a duality between land used with appropriate property titles and leases and squatted land. Recent empirical research has focused on the effects of the lack of effective, formal property titles which could prevent residents of squatter settlements from using their house as collateral. Informal land markets may thus be a major barrier to enterprise development. The empirical evidence about the relaxation of credit constraints associated with "titling" policies is weak. Recent work points instead to increases in labor supply (Field 2007) and to the adoption of more middle-class values and attitudes (Di Tella et al. 2007). While this evidence about titling policies is relatively optimistic about the merits of such policies, it must be noted that the existing literature focuses nearly exclusively on residential land. The extent of land illegality for commercial land (from illegal street vendors to squatter manufacturing) is poorly measured, and the solutions are not well developed.

The third key difference between cities in developing and more advanced countries regards infrastructure, particularly the road infrastructure. Two strands of research need to be distinguished here. The first finds its roots in international trade and focuses on the estimation of the effect of "market potential" variables. The market potential of a city is usually computed as the sum of the income (or population) of other cities weighted by their inverse distance to the city under consideration. Assuming transportation costs and other trade frictions associated with distance, many models of international and interregional trade generate the prediction that a location's income will be determined by its market access (Krugman 1991). The literature offers strong empirical support regarding the importance of market access for cities in developing countries (Henderson 2005).

The second strand of literature focuses more closely on the effects of infrastructure. Baum-Snow's (2007) pioneering work finds that the construction of the interstate highway system was a major impetus behind the suburbanization of US cities. Duranton and Turner (2012) also find that more kilometers of interstate

highways in US metropolitan areas in the early 1980s led to faster population growth over the subsequent 20 years.

This type of approach is also being applied to developing countries. In a remarkable piece of work, Donaldson (2010) documents the effects of the construction of India's railroad network by its colonial power. He shows that railroads increased trade and reduced price differences across regions. Even more importantly railroads increased real incomes and welfare. To minimize identification problems, he compares the network that was built to other networks that were considered but never developed.

In line with some of the arguments advanced above about the importance of transportation infrastructure for the decentralization of manufacturing activity away from large metropolises, Baum-Snow et al. (2011) underscore the importance of railroads in the decentralization of manufacturing production in China.

Storeygard (2011) provides evidence about the importance of inter-city transportation costs for inland African cities. Using new roads data for Africa and satellite data ("lights at night") to estimate economic activity, he assesses the effect of higher transportation costs. To circumvent the endogeneity of transportation costs (roads may be built to access growing cities), he uses arguably exogenous variations in oil prices. He finds an elasticity of economic activity with respect to transportation costs of about -0.2.

All these findings are suggestive of the profound and long-lasting effects of major transportation infrastructure. One needs to keep in mind nonetheless that major transportation networks are extremely costly investments.

The last key difference between cities in developing and more advanced countries regards the effects of the favoritism by governments of the largest cities. While the reasons for primate city favoritism are still debated (Henderson 2005), there is little doubt that such favoritism takes place in many different ways. As argued in Henderson (2005), primate city favoritism harms the favored primate city by making it bigger than it should be. It also harms smaller cities which are, in effect, heavily taxed. The gap that is created between the primate city and other cities may also have negative dynamic effects since for most educated workers there is nowhere to go except stay in this primate city. As a result this may reduce the circulation of knowledge across cities. Reducing primate city favoritism and providing smaller cities with better local public goods (including education and health) are certainly a big part of any solution.

33.5 Conclusions

For individual workers, cities in developing countries appear to bring significant benefits both in the short run and in the long run. However, when taking a broader look, the urban system of developing countries appears to involve far less functional differentiation across cities than in more advanced economies. Such differentiation with different cities playing different roles in the urban system is important for the process of growth and development to proceed smoothly.

Larger cities innovate and manage. Smaller cities often produce a narrow range of goods. Having larger cities do everything like they often in developing countries reduces their dynamism and holds back small cities which remain stagnant.

A variety of policies can be envisioned to solve this problem. The three more promising areas are general policies to improve the functioning of labor markets, ending primate city favoritism, and development of major infrastructure to connect cities.

References

Bartelsman E, Haltiwanger J, Scarpetta S (2004) Microeconomic evidence of creative destruction in industrial and developing countries. University of Maryland, Mimeo

Baum-Snow N (2007) Did highways cause suburbanization? Q J Econ 122(2):775–805

Baum-Snow N, Brandt L, Henderson JV, Turner MA, Zhang Q (2011) Roads, railways and decentralization of Chinese cities. Brown University (Processed)

Carlino GA, Chatterjee S, Hunt RM (2007) Urban density and the rate of invention. J Urban Econ 61(3):389–419

Charlot S, Duranton G (2004) Communication externalities in cities. J Urban Econ 56(3):581–613

Ciccone A, Hall RE (1996) Productivity and the density of economic activity. Am Econ Rev 86(1):54–70

Cingano F, Schivardi F (2004) Identifying the sources of local productivity growth. J Eur Econ Assoc 2(4):720–742

Combes P-P, Duranton G, Gobillon L (2008) Spatial wage disparities: Sorting matters! J Urban Econ 63(2):723–742

Combes P-P, Duranton G, Gobillon L, Puga D, Roux S (2011) The productivity advantages of large cities: distinguishing agglomeration from firm selection. Econometrica 80(6):2543–2594

De la Roca J, Puga D (2012) The dynamic earnings premium of dense cities. CEMFI and IMDEA Sociale Sciences (Processed)

Di Tella R, Galliani S, Schargrodsky E (2007) The formation of beliefs: evidence from the allocation of land titles to squatters. Q J Econ 122(1):209–241

Donaldson D (2010) Railroads of the Raj: estimating the impact of transportation infrastructure. MIT (Processed)

Duranton G, Puga D (2001) Nursery cities: urban diversity, process innovation, and the life cycle of products. Am Econ Rev 91(5):1454–1477

Duranton G, Puga D (2005) From sectoral to functional urban specialisation. J Urban Econ 57(2):343–370

Duranton G, Turner MA (2012) Urban growth and transportation. R Econ Stud 79(4):1407–1440

Ellison G, Glaeser EL, Kerr WR (2010) What causes industry agglomeration? Evidence from coagglomeration patterns. Am Econ Rev 100(3):1195–1213

Feldman MP, Audretsch DB (1999) Innovation in cities: Science-based diversity, specialization and localized competition. Eur Econ Rev 43(2):409–429

Field E (2007) Entitled to work: urban property rights and labor supply in Peru. Q J Econ 122(4):1561–1602

Figueiredo O, Guimarães P, Woodward D (2002) Home-field advantage: location decisions of Portuguese entrepreneurs. J Urban Econ 52(2):341–361

Freedman M (2008) Job hopping, earnings dynamics, and industrial agglomeration in the software publishing industry. J Urban Econ 64(3):590–600

Fujita M, Ishii R (1998) Global location behavior and organizational dynamics of Japanese electronics firms and their impact on regional economies. In: Chandler AD Jr, Hagström P, Sölvell Ö (eds) The dynamic firm: the role of technology, strategy, organization and regions. Oxford University Press, Oxford, pp 343–383

Glaeser EL, Gottlieb JD (2008) The economics of place-making policies. Brook Pap Econ Act 1:155–253
Glaeser EL, Kerr WR (2009) Local industrial conditions and entrepreneurship: how much of the spatial distribution can we explain? J Econ Manag Strategy 18(3):623–663
Glaeser EL, Resseger MR (2010) The complementarity between cities and skills. J Reg Sci 50(1):221–244
Glaeser EL, Kallal H, Scheinkman JA, Shleifer A (1992) Growth in cities. J Polit Econ 100(6):1126–1152
Glaeser EL, Kerr WR, Ponzetto GAM (2010) Clusters of entrepreneurship. J Urban Econ 67(1):150–168
Greenstone M, Hornbeck R, Moretti E (2010) Identifying agglomeration spillovers: evidence from winners and losers of large plants openings. J Polit Econ 118(3):536–598
Henderson JV (2005) Urbanization and growth. In: Aghion P, Durlauf SN (eds) Handbook of economic growth, vol 1B. North-Holland, Amsterdam, pp 1543–1591
Hsieh C-T, Klenow PJ (2009) Misallocation and manufacturing TFP in China and India. Q J Econ 124(4):1403–1448
Kerr WR (2010) Breakthrough inventions and migrating clusters of innovation. J Urban Econ 67(1):46–60
Krugman PR (1991) Increasing returns and economic geography. J Polit Econ 99(3):484–499
Lin J (2011) Technological adaptation, cities, and new work. Rev Econ Stat 93(2):554–574
Michelacci C, Silva O (2007) Why so many local entrepreneurs? Rev Econ Stat 89(4):615–633
Moretti E (2004) Human capital externalities in cities. In: Henderson V, Thisse J-F (eds) Handbook of regional and urban economics, vol 4. North-Holland, Amsterdam, pp 2243–2291
Porter ME (1990) The competitive advantage of nations. Free Press, New York
Puga D (2010) The magnitude and causes of agglomeration economies. J Reg Sci 50(1):203–219
Storeygard A (2011) Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa. Brown University (Processed)
Wheeler CH (2008) Local market scale and the pattern of job changes among young men. Reg Sci Urban Econ 38(2):101–118