



3068

REPRINT

Jacques-François Thisse

Economics of agglomeration

*Oxford Research Encyclopedia, Economics and Finance,
Oxford University Press, 2019.*



CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

Email: lidam-library@uclouvain.be

[https://uclouvain.be/en/research-institutes/
lidam/core/reprints.html](https://uclouvain.be/en/research-institutes/lidam/core/reprints.html)

Economics of Agglomeration

Jacques-François Thisse

Subject: Urban, Rural, and Regional Economics Online Publication Date: Mar 2019

DOI: 10.1093/acrefore/9780190625979.013.152

Summary and Keywords

Despite the drop in transport and commuting costs since the mid-19th century, sizable and lasting differences across locations at very different spatial scales remain the most striking feature of the space-economy. The main challenges of the economics of agglomeration are therefore (a) to explain why people and economic activities are agglomerated in a few places and (b) to understand why some places fare better than others.

To meet these challenges, the usual route is to appeal to the fundamental trade-off between (internal and external) increasing returns and various mobility costs. This trade-off has a major implication for the organization of the space-economy: High transport and commuting costs foster the dispersion of economic activities, while strong increasing returns act as a strong agglomeration force.

The first issue is to explain the existence of large and persistent regional disparities within nations or continents. At that spatial scale, the mobility of commodities and production factors is critical. By combining new trade theories with the mobility of firms and workers, economic geography shows that a core periphery structure can emerge as a stable market outcome.

Second, at the urban scale, cities stem from the interplay between agglomeration and dispersion forces: The former explain why firms and consumers want to be close to each other whereas the latter put an upper limit on city sizes. Housing and commuting costs, which increase with population size, are the most natural candidates for the dispersion force. What generates agglomeration forces is less obvious. The literature on urban economics has highlighted the fact that urban size is the source of various benefits, which increase firm productivity and consumer welfare.

Within cities, agglomeration also occurs in the form of shopping districts where firms selling differentiated products congregate. Strategic location considerations and product differentiation play a central role in the emergence of commercial districts because firms compete with a small number of close retailers.

Keywords: agglomeration, city, commuting, land, location, monopolistic competition, region, spatial competition, trade, transport

Introduction

Economic activities are concentrated in a limited number of cities and regions while spatial inequality is present at all spatial (and even historical) scales. Twenty metropolitan areas produce around 50 % of the GDP in the United States but are home to approximately 30 % of the population. In 2006, the metropolitan areas of Budapest, Seoul, Copenhagen, Dublin, Helsinki, and Randstad-Holland produced nearly half of their respective countries' national GDP, while London, Paris, Prague, Stockholm, and Tokyo accounted for about one third. However, a significant number of these metropolitan regions have large and persistent pockets of poverty. At the regional level, within-country disparities explain a growing share of spatial inequality in GDP per capita in the European Union.

At first glance, the existence of a spiky economic space may come as a surprise. During the past two decades, the media have embraced the idea that we live in a world where distance no longer matters. The steady drop in transport costs since the mid-19th century—compounded by the near disappearance of communication costs—would have freed businesses and households from the need to be located near one another. In other words, the space-economy is no longer “spiky,” but “flat.”

But—and it is a big but—despite the fact that proximity to natural resources has declined significantly, distance and location still matter in economic life. Recent work in spatial economics shows that the space-economy is being shaped by new forces, which had previously been overshadowed by natural factors. The new world is anything but flat. Distance and location still matter because they affect the economic life under different guises and through different channels.

The main challenges of the economics of agglomeration are therefore (a) to explain why people and economic activities are agglomerated in a few places; (b) to understand why these places fare better than others; and (c) to figure out whether there is a relation of causality from one to the other. To meet this challenge, it is usual to appeal to the fundamental trade-off between increasing returns and transport costs.

The intuition behind this trade-off is easy to grasp. Under constant (or decreasing) returns and positive transport costs, firms can unfold their production over a large number of locations without any efficiency loss, thus bringing transport costs down to zero. Under increasing returns and costless transport, firms would concentrate their production within a few giant plants to benefit from the highest possible level of efficiency. Such consequences are nonsensical and confirm the importance of the trade-off between increasing returns and transport costs, which tells us something important about the organization of the space-economy: High transport costs promote the dispersion of economic activities, while strong increasing returns act as a strong agglomeration force.

The fundamental trade-off of spatial economics has another major implication: If many activities can be located almost anywhere, few activities are located everywhere. It is in this sense that location matters: Although a growing number of activities become footloose with the development of new transport and communication means, a relatively small num-

ber of places in many countries account for a large share of the national value added, whereas many large areas have no or little economic activity.

The concept of economic agglomeration refers to very distinct real-world situations, which run from the global and national to the urban and local. Economists and regional scientists have developed different models to account for the existence of spatial inequality at different geographical scales. The choice of a geographical scale depends on the question being addressed. However, the differences are less marked than what they may seem: A few basic ideas are sufficient to understand the reasons for the existence of a wide range of diversified clusters. These ideas will serve as the backbone of this article.

For decades, the major contributions to spatial economics have been painfully deficient. One possible explanation is that spatial economics is fraught with difficulties that have been put aside in general competitive analysis. This problem is discussed more fully in the section “THE MARKET STRUCTURE PROBLEM IN SPATIAL ECONOMICS.” Subsequent sections deal with topics that occupy center stage in contemporary spatial economics.

The first section addresses the regional question, that is, the existence of sizable and lasting disparities in GDP per capita and income within and between nations. To be precise, the section surveys what has been accomplished in economic geography ever since the pioneering work of Krugman (1991). What distinguishes economic geography from trade theory is the mobility of firms and households and the way this mobility interacts with the distribution of product demand and labor supply. Economic geography has given new life to the study of macro-regions, which since then has made enormous progress by any previous yardstick.

The second section then examines the following fundamental questions: (a) Why do (big) cities exist? (b) What are the reasons for urban land rents that often vastly exceed the agricultural rent? Urban economics, in the wake of its founding fathers (Alonso, 1964; Mills, 1967), has focused on the monocentric city model in which jobs are supposed to be concentrated in the city’s central business district (CBD). This model provides an answer to the second question but remains silent about why a CBD would exist. Interest has recently shifted toward the reasons for the agglomeration of firms and households in cities, characterized by one or several employment centers, and an urban wage premium, which rises with the size of cities. This body of literature sets the stage for answering the first question by highlighting the benefits associated with a high population density.

The last section studies how firms belonging to the same industry choose to establish themselves in particular places. More specifically, the reasons for the existence of shopping districts are examined, where several firms selling similar products, such as restaurants, movie theatres, or fashion clothes, locate together to attract geographically dispersed consumers.

This article thus covers the three main subfields of spatial economics, namely *economic geography*, *urban economics*, and *location theory*. It aims to provide a quick, though comprehensive overview, of these subfields. A brief discussion of a possible synthesis concludes the article.

The Market Structure Problem in Spatial Economics

To gain insights into the nature of the problem, the system of forces that push and pull firms and consumers across space is described in an admittedly crude, but relevant way. Firms and households benefit from greater proximity to one another but face tougher competition in the use of scarce resources, such as land. Consumers are attracted by places where the density of firms is high because opportunities there are more numerous; they are discomfited by places with a high density of consumers because they dislike congestion. Firms are attracted to locations where the density of consumers is high because, there, the expected volume of business is large; they reject places where the density of sellers is high because of the stronger competition. Therefore, production and consumption are locationally interdependent. Such interactions are foreign to the competitive framework.

The Spatial Impossibility Theorem

Land and transport are the two quintessential spatial commodities. So, the following questions naturally come to mind: Can land and transport be taken into account in the general equilibrium model? If yes, can the competitive price mechanism be used to explain the agglomeration of economic activities in a small number of places, as well as the existence of sizable disparities at different spatial scales?

The most elegant and general model of a competitive economy is that developed by Arrow and Debreu. In this model, a commodity is defined not only by its physical characteristics, but also by the location where it is made available. Given this convention, the choice of a commodity must also entail the choice of a location. For example, when individuals choose a consumption good or a type of work, they also choose where to consume or where to work. Within the Arrow-Debreu model, spatial interdependencies are integrated in the same way as other market interactions. In other words, this model seems able to cope with the formation of a space-economy without having to consider additional specificities.

General equilibrium theory is useful for the study of commodity flows across space when both firms and households have given locations because the transport of commodities is consistent with the standard assumptions of the competitive setting. To illustrate this, consider an economy with a finite number of locations connected by a transport network. The issue is to determine simultaneously the quantities supplied and demanded in each

location and the local prices at which the good is available at different locations. Since trade is driven by spatial arbitrage, the price of a good in one location in equilibrium depends on the price for the same good in other locations as arbitrage limits the price difference to the shipping cost of the good. To be precise, the equilibrium is reached when the consumer price equals the producer price plus the transport cost for all positive flows; if the consumer price is less than the producer price plus the transport cost, there is no trade between these two locations.

However, as shown by Starrett (1978) in a fundamental paper, things are not that simple when the problem is no longer with transporting goods, but rather with locating firms and households. Although firms may run several production plants, they typically concentrate production in a few sites. Likewise, a household lives in a very small number of places, often one. It is therefore convenient to assume that each firm (household) chooses a single location and engages in its production (consumption) activities therein. In other words, agents are not ubiquitous and have an address in space. Even though agents are free to choose the amount of land they consume, agents are indivisible. Since firms and households are free to choose their addresses, together with the amounts of land they use, they make mutually exclusive choices. Once they are located, firms and households buy and sell goods and services. To achieve their goal, they must bear transport costs, which correspond to the expenditures needed to conduct transactions with remote suppliers and customers.

To illustrate the nature of Starrett's (1978) result, consider an economy with two locations, A and B , two price-taking firms, two tradable goods, and one non-tradable good, land. The argument still holds in the case of several firms, locations, and goods. Firm i ($= 1, 2$) produces good i ($= 1, 2$) by using land and good $j \neq i$ as inputs. Assume for the moment that there is not enough land in any location for the two firms to locate side by side. Denote by $p_{i\ell}$ the price of good $i = 1, 2$ in location $\ell = A, B$, by R_ℓ the price of one unit of land in $\ell = A, B$, and by $t_i > 0$ the unit transport costs of good i between A and B .

Assume a competitive equilibrium exists, that is, a price system and a production plan for each firm which satisfy the following conditions: markets clear and each firm maximizes its profit subject to its production function. In particular, no firm wants to change its location. Assume also that Firm 1, say, is established in A , while Firm 2 is set up in B . Since each firm sells its output to the other firm, the equilibrium prices of the goods must satisfy the following conditions:

$$p_{1B} = p_{1A} + t_1 > p_{1A}, \quad p_{2A} = p_{2B} + t_2 > p_{2B}.$$

(1)

The expression (1) means that the price of good 1 (2) in the destination B (A) is equal to its price in the origin A (B) plus the unit transport cost $t_1 > 0$ ($t_2 > 0$) of this good between the two locations. The inequalities are strict because shipping goods 1 and 2 between locations A and B is costly.

In equilibrium, Firm 1 maximizes profit by producing the quantity $\bar{y}_1$ of good 1 and by using $\bar{y}_2$ units of good 2 and $\bar{s}_1$ units of land. Its maximum profit is then given by

$$\pi_{1A} = p_{1A}\bar{y}_1 - p_{2A}\bar{y}_2 - R_A\bar{s}_1.$$

Since firms are price-takers, they believe that “changing place” does not affect the prevailing prices of goods and land rents. Therefore, if Firm 1 were established in B rather than A , the profit generated by the same production plan $(\bar{y}_1, \bar{y}_2, \bar{s})$ would be given by the following expression:

$$\pi_{1B} = p_{1B}\bar{y}_1 - p_{2B}\bar{y}_2 - R_B\bar{s}_1.$$

Using expression 1, it is readily verified that

$$\pi_{1B} - \pi_{1A} = t_1\bar{y}_1 + t_2\bar{y}_2 + (R_A - R_B)\bar{s}_1.$$

(2)

By symmetry, we have

$$\pi_{2A} - \pi_{2B} = t_2\hat{y}_2 + t_1\hat{y}_1 + (R_B - R_A)\hat{s}_2.$$

(3)

Summing Equations 2 and 3 and using expression 1 yields the following expression:

$$\mathbf{I} \equiv \pi_{1B} - \pi_{1A} + \pi_{2A} - \pi_{2B} = t_1(\bar{y}_1 + \hat{y}_1) + t_2(\bar{y}_2 + \hat{y}_2) + (R_A - R_B)(\bar{s}_1 - \hat{s}_2).$$

Without loss of generality, we can rename the two locations for $(R_A - R_B)(\bar{s}_1 - \hat{s}_2)$ to be nonnegative. In this case, the magnitude $\mathbf{I}$ is strictly positive because $t_i > 0$. Consequently, Firm 1 would be better off by moving to location B . Since the components of the vector $(p_{1A}, p_{1B}, p_{2A}, p_{2B}, t_1, t_2, R_A, R_B)$ are arbitrary, this means that there exists no competitive equilibrium in which the two firms ship their outputs between A and B .

Obviously, the two firms locate together if one location is endowed with enough land to accommodate the two firms. Moreover, in the absence of transport costs ($t_1 = t_2 = 0$), the aggregate incentive to move is equal to zero ($\mathbf{I} = 0$). Hence, what prevents a competitive equilibrium with trade from existing is the presence of transport costs, without which there is absolutely no role for geography, combined with the fact that firms have a specific address in space.

Furthermore, in the example, it is supposed that the production plan $(\bar{y}_1, \bar{y}_2, \bar{s})$ is feasible in both locations. More generally, the location space is said to be homogeneous if (a) the utility function of a household is the same no matter what its location and (b) the production function of a firm is independent of its location.

The argument can be extended to cope with households whose incentive to move is measured by the difference in expenditures while consuming the same bundle of goods. Summing across all agents and locations and using the material balance conditions shows that the aggregate incentive $\mathbf{I}$ to move exceeds carriers' costs, and thus is strictly positive. Since the aggregate $\mathbf{I}$ must be non-positive for an equilibrium to exist, one may conclude as follows:

The spatial impossibility theorem: Assume a finite and homogeneous location space. If firms and households consume land and if transport is costly, no feasible location pattern of firms and households with trade can be sustained as a competitive equilibrium.

When locations are not in autarky, some goods are traded across locations. By implication, the price system must perform two different jobs simultaneously: (a) support the shipping of goods between locations while clearing markets in each location, and (b) prevent firms and consumers from relocating. What the spatial impossibility theorem says is that the equilibrium prices supporting trade send the wrong signals from the viewpoint of locational stability. This is exactly what the 2-firm-2-location example states. Note that the spatial impossibility theorem uses the tools of general equilibrium theory itself to show the inability of the latter theory to explain the main features of a space-economy.

Comparative Advantage

One way to rescue the competitive paradigm in spatial economics is to relax the assumption of a homogeneous space and to appeal to comparative advantage. To see how this works, consider the previous 2-firm-2-location example where location A is now endowed with an exogenous attribute beneficial only to Firm 1. Firm 2 experiences a similar advantage in sole location B . Whatever the reason, this attribute gives rise to additional earnings equal to $b_A > 0$ when Firm 1 is at A . Under these circumstances, Firm 1's profit at location A is as follows:

$$\pi_{1A} = b_A + p_{1A}q_1 - p_{2A}q_2 - R_A,$$

whereas π_{1B} is unchanged. Once again measuring the incentive to move from A to B by the difference in profit at A and B , one obtains:

$$\pi_{1B} - \pi_{1A} = t_1q_1 + t_2q_2 + R_A - R_B - b_A,$$

which is always negative when b_A is sufficiently large. Since the same holds for Firm 2, one may conclude that a competitive equilibrium involving trade may exist when firms have strongly diverging preferences for location attributes. However, firms tend to share similar preferences about locations because they tend to locate together.

Models of comparative advantage typically focus on the economic consequences of differences in productivity, the uneven distribution of immobile resources (natural harbors), and amenities that attract or repel firms and workers (climate). Spatial inhomogeneities

may lead to specialization and trade, very much like different endowments of production factors in the Heckscher–Ohlin theory of trade. One extreme form of comparative advantage is the Armstrong assumption used in the trade literature. Applied to the foregoing example, this assumption means that good 1 can be produced in location *A* and good 2 in location *B* only. In this case, a competitive equilibrium exists. However, it is not clear why good 1 and good 2 are to be produced in particular locations.

Spatial inhomogeneities may provide agents with the signal that is missing in a homogeneous space. An illustrative and powerful example of this is provided by the canonical model of urban economics. Insofar as jobs are concentrated in the CBD, the interdependence among workers is replaced by their sole accessibility to the CBD whose impact on workers can be captured by a competitive land rent market. In this setup, the competitive paradigm permits the development of a rich analysis of the structure of a monocentric city. However, the model fails to explain why a CBD exists. This problem is discussed further in the section “URBAN ECONOMICS.”

Although the existence of comparative advantage and spatial inhomogeneities is obviously pertinent, relying on them to explain the existence of large urban agglomerations and large trade flows amounts to playing Hamlet without the Prince. Indeed, one wants to understand why some places are more productive than others rather than assuming the existence of such differences. That said, the spatial impossibility theorem has a major implication for spatial economics: to study the existence of different types of economic agglomerations, one must consider externalities, increasing returns, or imperfect competition. By implication, the market, left to itself, delivers a suboptimal outcome.

Economic Geography

Many countries are characterized by large variations in population size, average income, regional production structure, cost of living, and distribution of jobs across regions. All these magnitudes are endogenous, and the values they take are determined by the interaction among markets, public policies, and the mobility of goods and production factors.

The idea of spatial interaction is central to economic geography. Broadly defined, spatial interaction refers to a wide array of flows subject to various types of spatial frictions, such as transport, communication, and commuting costs. Examples of these flows include traded goods, migration, capital, as well as the interregional transmission of knowledge and business cycle effects. The bulk of economic geography has been restricted to the interregional and international mobility of commodities and production factors, with a special emphasis on capital and labor.

Thanks to the appearance of the new trade theories, the economics profession became better equipped to understand the uneven development of regions. More specifically, economic geography relies on the idea that the fate of a region depends on the way it interacts with other regions. Admittedly, most of the main ideas have been around for a long

time. However, Krugman's economic geography has the fundamental merit of having framed most of those ideas within a general equilibrium model.

The Backbone of Economic Geography

Firms operate under internal increasing returns and monopolistic competition on the product market. Internal increasing returns are common in sectors where costly, efficiency-raising investments lead firms to concentrate production in a small number of plants. It is convenient to think of these firms as producing manufactured goods (or tradable services). Furthermore, the gravity law shows that distance and borders remain strong impediments to trade. This has belatedly led trade theorists to take transport costs into account in their setups. All of this has been accomplished by combining the Dixit-Stiglitz model of monopolistic competition and the iceberg transport technology. In this model, a large number of firms produce a manufactured good in the form of horizontally differentiated varieties sold to consumers who have a constant elasticity of substitution (CES) preference for variety:

$$U = \left(\int_0^n x_i^{\frac{\sigma-1}{\sigma}} di \right)^{\frac{\sigma}{\sigma-1}} \\ = xn^{\frac{\sigma}{\sigma-1}} \text{ when } x_i = x \text{ for all } i \in [0, n],$$

where $\sigma > 1$ is the elasticity of substitution between any two varieties and n the mass of varieties. Each variety is produced by a single firm and each firm produces a single variety using a fixed and marginal requirement of labor. By using a continuum of firms, economic geography neglects the strategic interactions that lie at the heart of the spatial competition models. The manufactured good is tradable. Under the iceberg transport technology, for one unit of the good to arrive at destination, $\tau > 1$ units must be shipped, the share $\tau - 1$ having melted on the way. This ingenious modeling trick allows the integration of positive shipping costs without having to deal explicitly with a transport sector.

To see how transport costs affect firm demands, consider an economy with firms located in region $r = A, B$, which set the same price p_r . The CES demand for a variety produced in region r is given by the sum of the demand in r and the demand stemming from region $s \neq r$:

$$q_r = p_r^{-\sigma} (Y_r P_r^{\sigma-1} + \phi Y_s P_s^{\sigma-1})^{-1/(\sigma-1)},$$

(4)

where Y_r is the total income in region r , P_r the price index that prevails in this region, that is,

$$P_r \equiv \left(n_r p_r^{-(\sigma-1)} + \phi n_s p_s^{-(\sigma-1)} \right)^{-1/(\sigma-1)},$$

(5)

where n_r is the number of varieties produced in region r , while $\phi \equiv \tau^{-(\sigma-1)}$. Clearly, the higher the income of a region, the higher the demands for varieties. However, the size of the distant region is discounted by the factor ϕ that measures the freeness of trade: The higher the transport cost between A to B , the lower the actual demand from region B 's consumers a firm located in region A faces. In other words, for a firm located in region A , what matters is not the income Y_s of region s , but the spatially discounted income ϕY_s . Similarly, a higher number of competitors located in region s leads to a lower demand for a variety produced in region r . However, this number is also weighted by the spatial discount factor ϕ because firms established in a distant region affect demand less than do firms located in the same region. Therefore, everything seems to work as if the effective number of competitors, given by ϕn_s , would increase or decrease with ϕ . In sum, Equations 4 and 5 capture both the market size and the competitiveness of region r , as well as those of the other regions discounted by their corresponding spatial friction factor.

The combination of increasing returns and transport costs has a far-reaching implication: The proximity of large markets allows firms to better exploit the advantage generated by their own size and to save on the transport costs borne when supplying their customers. It is therefore no surprise that one of the main findings of economic geography is that firms are attracted by large markets. This is the main message of Krugman (1980): Firms located in big markets make higher profits, which in turn invite more firms to locate therein. Nevertheless, not all firms choose to locate in big markets. As firms set up in the large regions, competition is also heightened, thereby holding back the tendency toward agglomeration. Consequently, two forces pulling in opposite directions govern the interregional distribution of firms: Firms' desire for market access is the agglomeration force, while the dispersion force stems from firms' desire to avoid the presence of close competitors. This is the *proximity-competition trade-off*.

In the standard two-region setting, the larger market hosts a more than proportionate share of firms, a result known as the *home-market effect*. In other words, any initial advantage in market size is amplified by the home-market effect. Furthermore, wages paid in the larger region are higher than those in the smaller region (Takahashi et al., 2013). This result is in line with a major result discussed in the section "THE CORE-PERIPHERY STRUCTURE AS A STABLE MARKET EQUILIBRIUM," namely, wages are higher in larger cities. It also has an unexpected implication. Since the GDP per capita is higher in the larger region, capital flows from the poor region to the rich one, unlike what the standard constant returns-perfect competition paradigm predicts.

Unfortunately, the home-market effect cannot readily be extended to multiregional setups. Indeed, working with several regions brings about a new fundamental ingredient—the variability in regions' accessibility to spatially dispersed markets. In other words, the relative position of a region within the network of exchanges matters for this region's at-

tractiveness. Any global or local change in transport costs, such as those associated with a deeper market integration or the construction of a major transport link, triggers general equilibrium effects that vary in nontrivial ways with the shape and structure of the transport network.

Nevertheless, there is a wealth of empirical evidence suggesting that market access is associated with firms' location, higher wages, and low unemployment. In his survey of the literature, Redding (2011) concludes that "there is not only an association but also a causal relationship between market access and the spatial distribution of economic activity." In the same vein, Head and Mayer (2011) summarize their analysis of the relationship between market proximity and economic development over the 1965–2003 period by saying "market potential is a powerful driver of increases in income per capita." These results should not come as a surprise as the gravity equation stresses the importance of being close to rich areas for local firms to export.

The Core-Periphery Structure as a Stable Market Equilibrium

While the home-market effect may help in understanding how an initial advantage in market size may be magnified, one must still explain why one regional market may become bigger than the other.

The most natural way to think of an agglomeration is to start with a symmetric and stable world and to consider the emergence of agglomeration as the outcome of a symmetry-breaking mechanism. It was not until Krugman (1991) that a full-fledged general equilibrium setting was proposed. More specifically, Krugman identified the conditions under which a symmetric distribution of producers and consumers becomes an unstable equilibrium in a world that is otherwise symmetric (see also Fujita et al., 1999).

The workhorse of the core-periphery (CP) model is again the Dixit-Stiglitz-iceberg setup. What distinguishes the CP model from standard trade theory is that consumers are geographically mobile. When individuals move to a new region, they bring with them both their production and consumption capabilities. In other words, they spend their income in the region where they settle. As a result, the migration of individuals changes the size of labor and product markets in the origin and in the destination regions. Migration thus generates additional effects that are disregarded in the proximity-competition trade-off.

Two effects are intertwined, one involving firms and the other workers. First, the increase in the number of workers in region *A*, and therefore of consumers, boosts local demand for the manufactured good, inducing more firms to locate in *A*. Because firms produce under increasing returns, operating in a market with a growing number of workers allows them to produce at a more efficient scale because the local market gets bigger. Furthermore, the home-market effect implies that an increase in the size of the larger market at the expense of the smaller market generates a more than proportional increase

in the share of the manufacturing sector established in region *A*. This pushes nominal wages upward.

Second, if the number of firms located in *A* increases, the number of locally produced varieties also increases and, consequently, the price of the tradable good decreases in this region. The two effects, in turn, spark an increase in real wages and thus a new flow of workers from region *B* to region *A* where, all else being equal, they enjoy a higher standard of living. Therefore, there is cumulative causality as these two effects reinforce each other. This snowball effect seems to lead inevitably to the agglomeration of firms and workers in one region, which becomes the core of the economy.

To avoid this difficulty, Krugman added a genuine, but ad hoc dispersion force to his set-up by considering a second sector (e.g., agriculture). Farmers are spatially immobile and evenly distributed between the two regions. The agricultural good is produced under constant returns and perfect competition. Assuming that this good is traded at no cost implies that it is sold at the same price in the two regions. Consequently, farmers earn the same income in both regions and have the same demand for the manufactured good, which is rooted in the region where they live. This dispersion of demand incites manufacturers to choose different locations because they enjoy a proximity advantage in supplying local farmers.

The great accomplishment of Krugman (1991) was to integrate all these effects within a single framework and to determine the precise conditions under which the agglomeration or dispersion of firms and workers occurs. Turning next to the specific conditions for agglomeration, Krugman has shown that the value of transport costs is the key determining factor.

The timing of events is as follows: First, workers choose their locations. Second, given the interregional population distribution, production takes place. If transport costs are sufficiently high, interregional shipments of goods are expensive, which strengthens the dispersion force. Under these circumstances, the economy displays a symmetric regional pattern in which firms focus on local markets. By contrast, when transport costs are sufficiently low, this pattern ceases to be stable, and thus there is symmetry-breaking. In this case, all manufacturers concentrate in a single region, which becomes the core of the economy, whereas the other region, which supplies only the agricultural good, becomes the periphery. In this configuration, manufacturers are able to exploit increasing returns by selling more in the larger market without losing much business in the periphery, since transport costs are low while demand has shrunk due to the outmigration of workers.

The CP model thus allows for the possibility of convergence or divergence between regions, whereas the neoclassical model, based on constant returns and perfect competition, would predict convergence only. Consequently, it is fair to say that Krugman has presented a synthesis of the polarization and standard neoclassical theories. It is worth stressing here that the emergence of a CP structure is the involuntary consequence of decisions made by a large number of agents pursuing their own interests.

To sum up, we have the following result:

The core-periphery structure: When transport costs are sufficiently low, the manufacturing sector is agglomerated in a single region. Otherwise, this sector is equally dispersed between the two regions.

Workers' nominal wage is the same under agglomeration and dispersion, which seems weird. However, their real wage is higher under the former than the latter and, therefore, they are better off at the equilibrium featuring agglomeration.

Pros and Cons of the Core-Periphery Model

(1) The dimensionality problem mentioned in the study of the home-market effect also occurs in the CP model. Akamatsu et al. (2012) revisited the CP model in the case where 2^n regions are equidistantly distributed around a circle. Transport costs between any two regions thus vary with these regions' relative positions on the unit circle. Starting from a transport cost value that is high enough for the uniform distribution of the manufacturing sector to be a stable equilibrium, a gradual decrease in transport costs leads to a pattern in which workers and firms are concentrated in 2^{n-1} regions separated by the distance $1/2^{n-1}$. As transport costs keep decreasing, the CP model displays a cascade of bifurcations in which the number of manufacturing regions is reduced by half and the spacing between each pair of neighboring manufacturing regions doubles after each bifurcation. In the end, when transport costs are very low, the manufacturing sector is agglomerated into a single region, as in Krugman (1991).

Thus, the main conclusion drawn from the CP model does not hinge on the two-region setting used by Krugman. However, the set of possible configurations is much richer than in the original CP model: One observes simultaneously partial agglomeration and partial dispersion; that is, manufacturers are located in a few regions while the other regions do not host any. Note that regions that host manufacturers have the same size. How asymmetric regions may emerge is still under scrutiny.

(2) Despite the extreme simplicity of the CP model, its welfare analysis does not deliver an unambiguous message. When individuals move, they do not account for all the effects caused by their decisions. This implies the existence of pecuniary externalities, which find their origin in the fact that prices do not reflect the social value of individual decisions, since there is imperfect competition. As a result, the equilibrium is suboptimal. However, the inefficiency of the market outcome does not say anything about the excessive or insufficient concentration of activities in the core. Neither agglomeration nor dispersion Pareto-dominates the other. Indeed, workers and farmers living in the core always prefer agglomeration because all varieties are produced where they reside. By contrast, farmers living in the periphery always prefer dispersion because some varieties are produced therein. In other words, the shift from dispersion to agglomeration generates both welfare gains and welfare losses.

In order to evaluate the social desirability of agglomeration, one can appeal to compensations paid either by those who gain from the move or by those who are hurt by the move, using the corresponding market prices and wages (Charlot et al., 2006). When transport costs are sufficiently low, those located in the core can compensate those located in the periphery to sustain the utility level they enjoyed under dispersion. Indeed, firms' efficiency gains are high enough to overcome the losses incurred by the peripheral farmers. In this case, regional disparities are the geographical counterpart of a greater productive efficiency. However, when transport costs take on intermediate values, no clear recommendation emerges. This lack of sharp results may explain why so many divergent views coexist in a domain where there are good reasons to believe that different tenets are correct.

(3) The sudden and discontinuous shift from dispersion to agglomeration is a byproduct of the assumption of homogeneous workers—very much like in Bertrand's duopoly model of industrial organization. Yet it is well documented that potential migrants account for a wide range of non-pecuniary factors that act as an impediment to migration. Once it is recognized that individuals are heterogeneous because they have different attitudes toward the non-monetary attributes, the agglomeration process is gradual and sluggish. More importantly, migration costs are a dispersion force because they incite workers to stay put. This implies that there is no longer any need to make the ad hoc assumption of an immobile agricultural sector. Revisiting the CP model with imperfectly mobile workers, Murata (2003) shows that, when markets are sufficiently integrated for the real income gap to fall below the utility loss generated by homesickness, the agglomeration process is reversed. In sum, market integration fosters first divergence and then convergence. In other words, economic integration would yield a bell-shaped curve of spatial development.

(4) Studying the interaction between economic growth and the location of economic activity is a hard task. However, it is possible to derive some interesting results by adding to the CP model a research and development (R&D) sector. This sector uses skilled labor to create the blueprints necessary to produce the new varieties. The fixed cost borne by a firm producing a variety is then equal to the cost of acquiring the corresponding patent. The corresponding model combines the demand effect generated by the migration of skilled workers and the productivity effect generated by the existence of spillovers among skilled workers.

In the spirit of the CP model, Fujita and Thisse (2003) assume that the skilled workers are mobile and produce patents only. Skilled workers are forward looking and benefit from knowledge spillovers. At a spatial equilibrium, the distribution of activities does not change while the total number of patents, varieties, and firms grows at the same constant rate. The presence of an R&D sector reinforces the cumulative causality that lies at the heart of the CP model. In particular, the growth rate of the economy varies with the spatial distribution of skilled workers: It takes on its highest value when the R&D sector is agglomerated.

The additional growth spurred by agglomeration may lead to a Pareto-dominant outcome (i.e., when the growth effect triggered by agglomeration is strong enough, even those who stay put in the periphery are better off than under dispersion). Although the two regions grow at the same rate, the income gap widens: The rich get richer and so do the poor, but without ever catching up. How to make a decision in such a context is somewhat reminiscent of the ultimatum game. One possible way out of this dilemma is to influence the spatial extent of the spillover, since the growth of the economy depends on the spatial organization of the innovation sector across regions.

Beyond the Core-Periphery Model

To understand why large industrial regions develop in economies with a low spatial mobility of labor, one needs to move beyond Krugman to seek other explanations. Furthermore, economic geography has focused almost exclusively on falling transport costs. So, one may wonder about the spatial implications of technological progress in other sectors.

(1) The demand for intermediate goods is often greater than the demand for consumer goods. So final-goods producers, when choosing where to locate, pay attention to where intermediate-goods producers are situated; the reverse is also true. Therefore, if many final-goods producers are concentrated in one region and the demand for intermediate goods is high, producers of intermediate goods will be attracted to the area. With these producers close to the final-goods producers, prices of intermediate goods will be lower. This then attracts more final-goods producers to the region, resulting in a virtuous circle. Such a cumulative causality process feeds on itself, thus implying that agglomeration can be explained solely by the demand for intermediate goods, without having recourse to labor mobility, as in the CP model.

Giving intermediate goods a prominent role highlights forces, absent from the CP model, that are at work in modern economies. Once workers are immobile, a higher concentration of businesses within a region translates to a hike in local wages. This gives rise to two opposite forces. On the one hand, final demand in the core region increases because consumers enjoy higher incomes. Hence, final demand acts again as an agglomeration force. However, this force is no longer sparked by an increase in population size, but by an increase in income. On the other hand, an increase in wage generates a new dispersion force, which lies at the heart of many debates regarding the deindustrialization of developed countries; that is, their high labor costs. In such a context, firms are induced to relocate their activities to the periphery when lower wages there more than offset lower demand (Krugman & Venables, 1995). Thus, as transport costs fall there is, first, agglomeration and then redispatch of production.

In sum, economic integration would yield a bell-shaped curve of spatial development, which describes a rise in regional disparities in the early stages of the development process and a fall in later stages. Such a curve may be obtained in several extensions of

the CP model; it also seems to be confirmed by historical studies (Combes et al., 2011, and references therein).

(2) The CP model overlooks another fundamental fact; that is, firms are packets of functions, such as management, R&D, finance, and production. Due to the development of new information and communication technologies (ICTs), firms are able to disperse these functions into geographically separated units in order to benefit from the attributes specific to different locations. In this context, geographical separation generates a second type of spatial friction, namely communication costs. Coordinating activities within the firm is more costly when the headquarters and its production plants are physically separated because production requires both embodied and disembodied knowledge. If the former can be transmitted by using ICTs, the latter still requires face-to-face contacts because such contacts allow for immediate feedback and the utilization of tacit knowledge. In sum, lower communication costs make the coordination between headquarters and plants easier and therefore fosters the spatial fragmentation of firms. However, this claim requires qualification. For example, Henderson and Ono (2008) observe that the locations of U.S. manufacturing firms' headquarters is governed by a trade-off between the proximity to their plants and agglomeration economies available in large cities (see the section "URBAN ECONOMICS").

When "distance" is measured in two different ways, that is, by transport and communication costs, the intra-firm coordination costs must be sufficiently low for a plant to operate at a distance, whereas transport costs must decrease substantially to permit the supply of large markets at a low delivery cost from distant locations. When communication costs are high, all plants are located together with their headquarters and remain agglomerated in the core region, as in the CP model. However, once communication costs steadily decrease while transport costs are low, the industry moves toward a spatial pattern in which some firms are fragmented, whereas others remain integrated. Eventually, when communication costs have reached a sufficiently low level, the economy ends up with a deindustrialized core that retains only the firms' strategic functions, while what used to be the periphery accommodates a large share of the manufacturing sector. Thus, combining drops in transport and communication costs fosters convergence through the relocation of manufacturing plants from the developed to the developing countries (Fujita & Thisse, 2006).

This concurs with Baldwin (2016), who argues that there is no reason to expect that the ongoing ICT revolution will have the same effects on the space-economy as did the transport revolution that started in the mid-19th century. Instead, drops in transport and communication costs would foster the convergence between developed and developing countries.

(3) It is well documented that the transport sector has faced huge productivity gains during the past two centuries. However, a great number of other sectors have also experienced spectacular productivity gains. It is thus legitimate to ask what the impact is of the location of activity. When labor productivity grows in the manufacturing sector, regard-

less of its location, the mechanism at work in the CP model may be used to show that the most mobile workers, or the most productive firms, move to the larger region (Tabuchi et al., 2018). When productivity in manufacturing grows, fewer workers are needed to produce the differentiated product. Thus, workers become available to produce a greater output of the existing varieties or to work in new firms that provide new varieties; these effects will be stronger in the larger region than in the smaller. Consequently, a steady flow of innovations brings about a gradual transition from a dispersed configuration of the manufacturing sector to a partially agglomerated one. Rising labor productivity widens the real wage gap, which eventually outweighs some workers' migration costs and generates interregional migration. Hence, technological progress in manufacturing is a strong agglomeration force that owes little to falling transport costs.

Workers are also heterogeneous in skills. Those who move from *B* to *A* are often region *B*'s most skilled workers. Indeed, the skilled workers will benefit most from the hike in the price of one efficiency unit of labor. They also have the lowest mobility costs because their high level of human capital allows them to reduce the transaction costs generated by the lack of information they face in the region of destination. This affects the two regions in opposite ways: Region *A* tends to become more productive, whereas region *B* loses its best workers. As migrants get absorbed by the labor market of the core region, the agglomeration economies (discussed in the section "URBAN ECONOMICS") come into play, which reduces the number of job seekers and raises the productivity of labor in the region of destination. In addition, the arrival of skilled workers increases the incentives for the local unskilled to be trained while making education less attractive in the other region because it has already lost its best workers (Miyagiwa, 1991). In this context, interregional income and welfare gaps stem from the uneven geographical distribution of human capital, as confirmed by Combes et al. (2008) and Moretti (2012). Asserting how good or bad migrations are in dismantling regional disparities thus remains an open question.

(4) The mobility of workers is driven to a significant extent by the difference in real wages, which depend on the regional costs of living. Housing and commuting account for a large share in consumers' expenditures, but these costs are neglected in the CP model. In this respect, it is worth stressing here the work of Helpman (1998), who argues that decreasing transport costs may trigger the dispersion, rather than the agglomeration, of economic activities when the dispersion force lies in a perfectly inelastic housing supply, rather than immobile farmers. This difference in results is easy to understand. Housing costs rise when consumers move to the larger region, which strengthens the dispersion force, while lowering transport costs facilitates interregional trade, hence the convergence of commodity prices. In this case, the crowding of the housing market puts a brake on the agglomeration process, and thus Krugman's prediction is reversed. More generally, Murata and Thisse (2005) show that the interplay between transport and commuting costs is key in the formation of the space-economy.

By assuming that regions are dimensionless, the CP model remains in the tradition of trade theory. Economic geography may thus be viewed as spatial economics *without* land. Since the agglomeration of activities within a region often takes the form of cities, it is natural to turn one's attention to urban economics, which is spatial economics *with* land.

Urban Economics

Urbanization is probably the most extreme form of geographical unevenness, as the most distinctive feature of a city is its very high population density. This raises the following question: Why do consumers and firms seek physical proximity? Fundamentally, this proximity occurs because they want to interact. This need is gravitational in that its intensity increases with the number of agents set up in each location and decreases with distance, which has the nature of an impediment to interaction.

The industrial town is typically the location where production factors are combined under the same roof to take advantage of the division of labor. Today, the big city is the result of a richer causality, which includes the specialization and diversification of tasks provided by high-skilled workers; the widening of the range of choices of tradable consumption and intermediate goods and non-tradable services; the various types of communication externalities; and the provision of various local public goods. The usual cliché is that big cities are bad for consumers: Cities, especially those located in developing countries, are expensive, congested, polluted, and have slums and high crime rates. However, this argument forgets two things: (a) all over the world, people vote with their feet by moving to cities; and (b) cities are great places of consumption, culture, and leisure.

The Trade-off Between Housing and Commuting Costs

The monocentric city model is the canonical model of urban economics. It builds on von Thünen's work (1826/1966), whose purpose was to explain the pattern of agricultural activities surrounding cities in preindustrial Germany. Each location in space is characterized by various factors, such as soil conditions, relief, and geographical position. Among these factors, the most important for location theorists is the transport-cost differential over space. To this end, von Thünen used a very simple and elegant setting in which space is represented by a plain on which land is homogeneous in all respects except for a market town, where all transactions must occur. The location of the market town is supposed to be given and the reasons for its existence are left outside of the analysis.

By allocating an acre of land near the town for some specific crop, the costs of delivering all other crops are indirectly affected, as they must be grown farther away. von Thünen (1826/1966) very ingeniously imagined a process in which each farmer makes an offer based on the surplus he can generate by using one unit of land available at any particular location. This led von Thünen and his successors to develop the concept of bid rent func-

tion, which describes the maximum price a farmer is willing to pay to occupy each location. This price depends on the transportability of the output and the amount of land needed to produce one unit of the good. Land being allocated to the highest bidder, economic activities are distributed according to concentric rings, each specializing in one crop. The land rent decreases with distance from the market town at a rate that is constant in each ring; the land rent also decreases from one ring to the next.

Assuming the existence of a market town where transactions occur breaks down the locational interdependence among farmers. The only thing that matters to them is the distance to this town. Note also that assuming the existence of a market town whose location is given amounts to introducing a strong inhomogeneity in space that allows the use of perfect competition in the study of the land market. In large part, this explains why the monocentric city model has been, and still is, so popular.

Housing and commuting are the two main consumption items for households in developed countries. In the United States, the average expenditure share for housing is 24%, while it is 27% in France. Expenditures on transport amounts to 17% in the United States and 13.5% in France. Not surprisingly, therefore, the following two ideas lie at the heart of urban economics: (a) people prefer shorter trips to longer trips, and (b) people prefer having more space than less space.

Since activities cannot be concentrated on the head of a pin, they must be distributed across space. Alonso (1964) extended von Thünen's concept of bid rent curves to an urban context in which a market town is replaced by a CBD. In this context, the only spatial characteristic of a location is its distance from the city center, while the land available for raising crops is now used for housing. The main objective of the model is to explain the internal structure of cities and the land price gradient by focusing on households' trade-off between the desire to consume more space and accessibility to the CBD. In equilibrium, identical consumers establish themselves within the city so as to equalize utility across space. In this case, no one has an incentive to change location.

Consider a featureless plain with a dimensionless CBD located at $x = 0$ and a population of consumers who share the same income Y and the same preferences $U(z, s)$, where z is the consumption of a composite good, chosen as the numéraire, and s the amount of space used. Consumers compete to be as close as possible to the workplace. In any relatively big city, the amount of land available near the CBD is not sufficient to accommodate the entire population. So, how do consumers distribute themselves across the city? This is where the land market comes into play. The formal argument is disarmingly simple.

Denoting the land rent prevailing at a distance x from the CBD by $R(x)$ and the commuting cost borne by a consumer residing at x by $T(x)$, the budget constraint of this consumer is given by

$$z(x) + s(x)R(x) = I(x) \equiv Y - T(x).$$

Let $V(R(x), I(x))$ be the indirect utility of a consumer at x . Since the highest utility level attainable by consumers is the same across locations, the derivative of $V(R(x), I(x))$ with respect to x must be equal to zero:

$$\frac{\partial V}{\partial R} \frac{dR}{dx} + \frac{\partial V}{\partial I} \frac{dI}{dx} = 0.$$

Using Roy's identity and the equality $I_x = -T_x$, one obtains the Alonso-Muth condition:

$$s(x) \frac{dR}{dx} + \frac{dT}{dx} = 0.$$

(6)

Since a longer commute generates a higher cost ($T_x > 0$), this condition holds if and only if the land rent $R(x)$ is downward sloping ($dR/dx < 0$). The Alonso-Muth condition also shows that people choose to trade bigger plots for higher commuting costs. Indeed, Equation 6 means that a marginal increase in commuting costs associated with a longer trip is exactly compensated for by the income share saved on land consumption. Finally, as shown by Equation 6, if commuting costs were independent of distance ($dT/dx = 0$), the land rent would be constant and equal to the opportunity cost of land. Therefore, a high or low land rent reflects the good or bad access to the CBD (or to other places that attract people). In other words, commuting costs are the cause and land rents the consequence. More generally, the land rent prevailing at a particular location capitalizes—most of the time, imperfectly because of the land market failures—the costs and benefits of that location.

Furthermore, even though a longer commute is associated with a lower net income $Y - T(x)$, the lot size occupied by a consumer increases as the distance to the CBD rises. Indeed, the spatial equilibrium condition implies that the utility level is the same across all consumers. Therefore, the consumer optimization problem yields a compensated demand for land that depends on the land rent and the endogenous utility level. The utility level is treated as a given by consumers because every consumer is too small to affect it. Since housing is a normal good, a lower price for land therefore implies a higher land consumption. Consequently, as the distance to the CBD increases, the lot sizes increase whereas the consumption of the composite good decreases. By implication, the population density decreases as one moves away from the city center.

Though very simple, the monocentric city model has produced several results consistent with the prominent features of cities (Fujita, 1989; Duranton & Puga, 2015). For example, it explains how automobiles have generated suburbanization via lower unit commuting costs. However, very much as in the von Thünen model (which does not say why a given market town exists), the monocentric city model is silent on the reasons that would explain the existence of a district where jobs are available. So, one is left with the following question: Why do CBDs exist? Or, more generally, why do cities exist? A plausible answer

is that cities emerge as the outcome of a trade-off between agglomeration economies and urban costs. What, then, are these agglomeration economies?

Agglomeration Economies

The recent literature in urban economics explores the various external benefits arising when firms and consumers co-locate. Most importantly, they aim to explain why workers are better paid in larger cities (the urban wage premium). The bulk of the research has focused on business agglomeration economies, but there is a growing interest in the advantages associated with size for the urbanites. By crowding a few locations, firms and consumers reduce their land consumption. Therefore, one can view the formation of cities, at least in the first order, as the interplay between an interaction field among economic agents and competition on the land market; the former explains why firms and consumers want to be close to each other, whereas the latter puts an upper limit upon their spatial concentration.

Agglomeration Economies in Production

The main distinctive feature of external increasing returns is that they typically affect the firms belonging to the same small geographical area; thus, the name agglomeration economies. Such externalities do not spread over other regions or, to be more precise, their impact on distant regions is negligible. Since agglomeration economies are external to firms, all markets may be assumed to be perfectly competitive, thus providing a way out of the spatial impossibility theorem. Agglomeration economies endow cities, or at least some of them, with an endogenous comparative advantage. Agglomeration economies are often the unintentional outcome of a myriad of decisions made by economic agents pursuing their own interests, although public policies may help, or hurt, the unfolding of benefits associated with city size

Ever since Duranton and Puga (2004), it is customary to organize the various external effects that determine firms' productivity in cities into the following three classes: sharing, matching, and learning:

1. Sharing refers primarily to local public goods provided to consumers and producers. When seeking a reason for the existence of cities, the one that comes most naturally to mind is the variety and quality of public services, which help to enhance consumers' well-being. A large number of people facilitate the provision of local public goods that could hardly be obtained in isolation. One of the first historical examples is the walling in of cities, which exhibits increasing returns and has the nature of a local public good whose supply is influenced by size effects. The length of a circular wall is $2\pi\rho$, whereas the size of the corresponding area is $\pi\rho^2$. The ratio of the circumference to the area falls as the radius ρ increases, and thus a larger number of individuals may be defended at a lower average cost. In addition, firms' productivity is also positively affected by the presence of

some facilities, such as those required by the use of ICTs and various transport infrastructures.

However, sharing also refers to the supply of intermediate or business-to-business services, which are available in large cities. Even though firms outsource a growing number of activities to countries where labor is cheap, they also use a wider range of specialized services that are available only where these services are produced, that is, in big cities.

2. Matching means that the quality of matches between workers and firms on the labor market is higher in a thick market than in a thin one. Indeed, firms and workers operating in a denser labor market face a larger number of opportunities. Furthermore, sticky workers living in relatively small cities operate in markets with few potential employers, thereby allowing the local firms to exploit their monopsony power. By contrast, workers living in big cities face a wide range of job opportunities. This induces firms to pay higher wages because tougher competition on the labor market reduces their markdown.

3. Learning as an agglomeration force seems a priori less obvious than sharing and matching. It contends that different agents own different bits of information. So, getting the agents together allows knowledge spillovers that raise the overall level of knowledge, thus improving firms' and workers' productivity. Indeed, cities are the places where people talk.

Obviously, much of this talk does not generate productivity gains. However, the greater the number of people, the more likely the talk leads to new knowledge. The idea seems compelling: A greater number of agents owning different bits of information allows for the emergence of a wider range of combinations; this in turn fosters the emergence of a higher level of information and knowledge through various networks that make ideas available to more agents. That is precisely what cities do: *they connect people*. In this case, information has the nature of a spatial externality because, as it circulates within the local cluster of firms and workers, it contributes to aggregate productivity.

Since individuals learn from others, acquiring knowledge enhances not only the productivity of the worker who acquires it but also the productivity of others (Moretti, 2004). In addition, skilled workers seem to benefit from the presence of other skilled workers, more so than do unskilled workers. For example, Bacolod et al. (2009) observe that the urban wage premium associated with large cities stems more from cognitive skills than from motor skills. As a result, everything seems to work as if the productivity of a worker increases with the number of skilled workers working or living nearby. Clearly, this effect is stronger in the case of regular, easy contacts between skilled workers. It is no surprise, therefore, that workers with specific skills sort out across space according to their skills.

The empirical evidence seems to confirm this prediction. In the United States, Moretti (2012) observed that college graduates living in the richest cities, which are typically knowledge-based metropolitan areas, earn wages that are 50% higher than college graduates living in the bottom group of cities. In France, about half of spatial income disparities are explained by the different locations of skilled and unskilled workers (Combes et

al., 2008), while between 85% and 88% of spatial wage disparities in the United Kingdom are explained by individual characteristics (Gibbons et al., 2014). The concentration of human capital and high-value activities in large cities is a marked feature of developed and emerging economies. Therefore, spatial inequality tends more and more to reflect differences in the distribution of skills and human capital across space.

There is nothing new under the sun. In the 15th and 16th centuries, painting was one of the high-tech activities in Europe. In 1550, Giorgio Vasari wrote: "It is a habit of Nature when she makes one man very great in any art, not to make him alone, but at the same time and in the same place to produce another to rival him, that they may aid each other by emulation." The congruence of talent is rarely natural. Rather, it is generally due to the attractive power of prosperous cities (i.e., Venice, Florence, Antwerp, and Amsterdam in the past, and New York, London, Tokyo, and Shanghai in the 21st century). In other words, the geographical concentration of talents is anything but new. What is new is the growing importance of knowledge-based activities in the economy, hence the increasing need for highly skilled workers.

The downside of the spatial sorting of skilled workers across cities is the existence of stagnating or declining areas that specialized in industries associated with low wages and a small number of local consumer businesses. This decline may well have triggered strong and disruptive political effects because geographical social divides beget political divides (Moretti, 2012; Inglehart & Norris, 2016).

The existence of agglomeration economies is unquestionable (Combes et al., 2012; Combes & Gobillon, 2015). However, for both academic and policy reasons, it is important to go one step further and assess the specific sources of agglomeration economies. Measuring their relative magnitudes is hard. Ideally, one should test predictions that are associated with a single type of agglomeration economy, thus ruling out other sources; however, finding such data sets is difficult. A simpler research strategy is to consider the percentage change in productivity brought about by a higher employment (or population) density. An ordinary least squares regression of the logarithm of the average wage on the logarithm of the employment density across cities yields an elasticity that varies from 0.03 to 0.09 (Rosenthal & Strange, 2004).

However, estimations must be considered with care because some econometric problems may not have been properly addressed. In particular, workers are attracted by cities benefiting from wage increases and put off by those suffering negative shocks. Hence, one faces the problem of reverse causality: Wages would explain employment density, and not the other way around. As a result, employment density is necessarily correlated with the residuals, which contradicts one of the assumptions underpinning the validity of the ordinary least squares estimator. The most common solution involves using instrumental variable techniques.

In a comprehensive study, Faggio et al. (2017) confirm the presence of the various effects discussed, but stress the fact that agglomeration is a very heterogeneous phenomenon. For example, low-tech industries benefit from spillovers, but less than high-tech indus-

tries. Both intrasectoral and intersectoral external effects are at work, but they affect industries to a different degree. Firm size also matters: Agglomeration effects tend to be stronger when firms are smaller, perhaps because such firms do not benefit from the knowledge spillovers generated within firms that perform a large number of activities. In other words, specialized and vertically disintegrated firms would benefit more from spatial proximity than larger firms. Despite rapid progresses made since the late 1990s, more work is called for before having a deep understanding of the various effects at work.

The empirical evidence is compelling: Urban agglomeration economies are not an urban legend. That said, the following caution is in order: The literature on agglomeration economies, with its emphasis on the virtues of density, could convey the wrong impression that “big is beautiful.” Physical proximity and size are not enough to make a city affluent and successful. Cities are part of a whole, which includes national and local governments that are responsible for a wide range of public goods. Governance deficiencies, at any level of the political and administrative hierarchy, may prevent cities from developing their full potential. Many urban economists assume (implicitly) that the various institutions that affect the working of cities operate as they should, while the application of bylaws, regulations, and norms allows markets to function correctly. However, there is no functioning city without a functioning state.

Several small or medium-sized cities are quite successful, especially where they benefit from a good political and business environment. Smaller cities are both necessary and inevitable, as firms with greater relative space requirements and lower relative human capital requirements seek out the most cost-effective locations (Henderson, 1997). When they combine cheaper labor and land with a business-friendly local culture, smaller cities have a cost advantage that offsets their lack of agglomeration economies.

Can Agglomeration Economies Generate Employment Centers?

Ogawa and Fujita (1980) were the first authors who developed a general equilibrium model where a particular externality—knowledge spillovers—connect firms through a field of non-market interactions, while markets for goods, labor, and land are perfectly competitive. When firms own different pieces of information, the benefits of communication increase as the number of firms rises. Furthermore, since communications typically involve distance-decay effects, the benefits are greater if firms locate within the same district. This externality is the agglomeration force of their model.

The clustering of firms increases the average commute, which in turn leads workers to pay on average a higher land rent. Firms must therefore pay their workers a higher wage. In other words, the dispersion force stems from the interaction between the land and labor markets in firms’ optimization programs. The balance between those agglomeration and dispersion forces determines the equilibrium distribution of firms and workers. Note

the difference with the monocentric model in which the CBD is given: Here, the interactions among firms make the relative advantage of a given location for a firm dependent on the locations chosen by the others, which in turn affects workers' residential choices.

Ogawa and Fujita (1980) show that, in equilibrium, the city displays different configurations according to the level of commuting costs and the sensitivity of knowledge spillovers to distance. First, when commuting costs are high in relation to the distance-decay effect, the equilibrium involves a full integration of business and residential activities. To put it differently, land use is unspecialized. Since workers are located together with the firms where they work, there is what is called backyard capitalism. As commuting costs fall, two employment centers, which are themselves flanked by a residential area, are formed around an integrated area that involves firms and workers. Eventually, when commuting costs are low enough, the city becomes monocentric, with a CBD that now has a spatial extension. In this configuration, land use is fully specialized.

These results appear to concur with how cities organized spatially after the transportation revolution that followed the Industrial Revolution. When people moved on foot, activities were dispersed in preindustrial cities. With the industrial era, the large manufacturing CBD emerged. While modern cities still have large CBDs, factories with their need for land have been replaced by activities in land-intensive office space.

Since markets are perfectly competitive while the CBD is exogenously given, the market outcome is efficient in the monocentric city model (Fujita, 1989). However, in the presence of a spatial externality, this property ceases to hold. To be precise, firms care only about what they “receive” from the other firms and disregard the fact that they also “transmit” knowledge to other firms. As a result, the equilibrium is less concentrated than the optimum. Simply put, from the social standpoint, in the presence of spillovers, competition may well result in an insufficient concentration of population and activities within the city.

Despite the importance of the subject, only a limited number of papers have tackled the endogenous formation of employment centers. Duranton and Puga (2004) and Behrens and Robert-Nicoud (2015) provide detailed overviews.

Agglomeration Economies in Consumption

Cities are great places for consumers (Glaeser et al., 2001). Consumers living in large cities benefit from the same agglomeration economies as firms through a greater number of tradable and non-tradable goods and services, better transport and communication infrastructures, and a wider array of cultural amenities and opportunities for social relations. In addition, a larger city size allows more firms to enter the market, which leads to the provision of a greater number of available goods and services, higher quality, and more varieties of differentiated products. Tradable goods may also be cheaper in bigger cities (Handbury & Weinstein, 2015). By contrast, non-tradable goods and services tend

to be more expensive in larger cities because workers must be compensated for higher urban costs (Davis & Dingel, 2019).

The Social Costs of Urban Size

Big cities have their dark side. Housing is more expensive in big cities than in small cities. These high rents are often driven by excessive regulation of the housing and land markets. Various public policies often curtail the amount of land available for housing and offices, reducing the price elasticity of housing supply. Restrictive local planning hurts those who seek to move to the city or encourages those who do live in the city to move away. This prevents the most productive cities from fully benefiting from their potential agglomeration economies. In this case, restrictive land use regulations create an “artificial” scarcity of land that is reflected in the land rent level. Owners of existing buildings are often those who benefit from these restrictions (Glaeser & Gyourko, 2018).

Bigger cities also have much higher crime rates than smaller cities. Urban neighborhoods are often the substrate for the development of social norms such as conformity and status seeking, which govern the behavior of groups of individuals. Understanding why and how crime develops needs answering of the following difficult question: Why do people belonging to the same group or neighborhood tend to behave similarly? It is difficult to decipher the underlying mechanisms because herd behavior may arise for very different reasons (Manski, 2000): (a) the propensity of an individual to behave in some way depends on the behavior of the group; (b) the propensity of an individual to behave in some way depends on the exogenous characteristics of the group members; and (c) members of the group have similar individual characteristics and face a similar environment.

In addition, larger urban areas are on average more unequal and more segregated than smaller cities (Glaeser et al., 2008). Urban polarization is strengthened by spatial segregation in education and access to the job market. The mechanisms that create such educational and social traps display the same cumulative nature as those driving the formation of agglomeration economies, but here they generate vicious circles instead of virtuous ones. Owing to the cumulative nature of the mechanisms at work, policies aiming to promote urban cohesion should use powerful positive-discrimination instruments to improve education, health, housing, and safety. Preventing the development of a vicious circle is always much less costly than trying to cure it because vicious circles are difficult to reverse once set in motion. However, the necessary political will and popular support are often lacking when scarce resources must be allocated to solve a problem that seems small or temporary. Yet, when ghettos are formed, turning back is almost impossible.

Spatial Competition

Kaldor (1935) argued that space molds competition in a way that renders the hypothesis of perfect competition untenable. The argument goes as follows: When firms sell a homogeneous good, each consumer buys from the firm with the lowest full price that results

from adding the posted price to the travel cost of getting to the firm. Hence, each firm has some market power over the consumers located in its vicinity because buying from more distant firms is often more expensive. In other words, their geographical isolation provides them with only local monopoly power. As a result, whatever the total number of firms in the industry, each one competes more vigorously with its immediate neighbors than with more distant firms, thus implying that firms behave strategically in the sense of non-cooperative game theory. The global market is therefore segmented into several sub-markets formed by consumers who are more or less captive because their mobility is confined to an area determined by their residence and the firms they patronize. Such a market description is known as *spatial competition*.

Competition ceases to be monopolistic and becomes oligopolistic. One may think of the difference between the two settings as being the reflection of a difference in the spatial scale. The former (global) provides a good approximation of competition in the large, that is, competition among a large number of firms supplying different regions and countries; the latter (local) fits better as competition in the small, which involves a small number of firms located within the same city or neighborhood. However, one should keep in mind that location theory retains its relevance in the study of locational issues in various spatial environments.

The Hotelling Model

How do firms choose their locations under spatial competition? In a very influential paper, Hotelling (1929) proposed the following solution in the case of a spatial duopoly where the consumers live at locations continuously and uniformly distributed along a line segment (Main Street) and where two stores sell an identical good to dispersed consumers. The sellers choose, first, where to set up their stores along Main Street and then they choose the price at which they supply their customers. In game theory terms, the market outcome is thus given by the subgame perfect Nash equilibrium of a sequential game whose solution is obtained by backward induction: Firms find the Nash price equilibrium of the subgame induced by any location pair; using these prices, firms determine the Nash equilibrium of the location game.

For each pair of prices, the consumers partition their purchases between the firms by comparing the full prices. Each consumer buys one unit of the good at the firm with the lower full price. Consumers are thus divided into two segments, with each firm's aggregate demand represented by the consumers in one segment. The boundary between the two firms is given by the location of the consumer who is indifferent as to which firm is patronized. This boundary is endogenous because it depends on the prices set by the firms. Since the consumer distribution is continuous, a marginal variation in price leads to a marginal change in the boundary and to a change in demand of the same order.

According to Hotelling, at the equilibrium of the first-stage game, the two firms locate back-to-back at the market center. This provides a rationale for the agglomeration of firms selling to spatially dispersed customers. Unfortunately, Hotelling's analysis was

plagued by a mistake: When firms are sufficiently close, the corresponding subgame has no Nash equilibrium in pure strategies. This led d'Aspremont et al. (1979) to revisit the Hotelling setting by assuming that travel costs are quadratic in distance instead of linear. Such an assumption captures the idea that the marginal cost of time increases with the length of the trip. In this case, d'Aspremont et al. (1979) show that firms choose to set up at the two endpoints of Main Street. This extreme separation of firms arises because geographical separation relaxes price competition.

However, this negative result does not kill the subject. First, it confirms that competition is a strong dispersion force, as already highlighted by the proximity-competition trade-off of economic geography. Second, the standard model of spatial competition portrays consumers as robots that react to any small change in price or location. For example, if the two retailers are located side by side, a small price cut by one firm leads all consumers to shift to that firm. This extreme behavioral reaction seems unwarranted. The literature in marketing suggests that individual choices are better described by richer models, such as those discussed by McFadden (1986).

The Emergence of Shopping Districts

Building on this idea, de Palma et al. (1985) have reformulated the spatial competition model along the following lines: Firms now sell a differentiated product, while consumers' behavior is described by a purchasing probability (or frequency) that reflects their desire to try different varieties. However, this probability is endogenous, as it depends on the prices and locations selected by firms. More specifically, de Palma et al. (1985) model consumers' shopping behavior by using the multinomial logit (MNL): If Firm 2 is located at y_2 and sets the price p_2 , the probability that a consumer located at x buys from Firm 1 located at y_1 and selling at price p_1 is given by the following expression:

$$\mathbf{P}_1(x) = \frac{\exp[(p_1 + t|x - y_1|)/\alpha]}{\exp[(p_1 + t|x - y_1|)/\alpha] + \exp[(p_2 + t|x - y_2|)/\alpha]} > 0,$$

where parameter $\alpha > 0$ measures the degree of differentiation among varieties or the intensity of consumers' preference for variety. Note the similarity between the purchasing probability $\mathbf{P}_1$ associated with the MNL and the CES demand (Equation 4), which can be rewritten as follows in the case of two firms:

$$D_1(x) = \frac{p_1^{-1/\alpha}}{p_1^{-1/\alpha} + p_2^{-1/\alpha}} \frac{Y}{p_1},$$

with $1/\alpha = \sigma - 1$. This highlights the links between results obtained in economic geography and location theory.

The probability $\mathbf{P}_1$ rises when the price p_1 set by Firm 1 falls or when the distance $|x - y_1|$ between the consumer and Firm 1 decreases because this firm becomes more attractive.

The same holds when p_2 or $|x - y_2|$ increase. Furthermore, when consumers value more product variety, they are less sensitive to spatial proximity. In the limit, when $\alpha \rightarrow \infty$, distance becomes inessential and $\mathbf{P}_1(x) = 1/2$. However, when $\alpha = 0$ (the product is homogeneous), each consumer patronizes the cheapest retailer. In the Hotelling model, if the two firms are located side by side and charge identical prices, reversing their locations will exchange their clients. When individual behavior is probabilistic, turnover is not so abrupt. More specifically, if α is “sufficiently” large (see the discussion for what is meant by “sufficiently”), a unilateral reduction in a firm’s price (or a reversal of locations) no longer brings about the same upheaval in the allocation of customers between firms.

When the preference for variety is sufficiently strong and travel costs are sufficiently low for $\alpha > t/2$ to hold, de Palma et al. (1985) show that the two firms locate at the market center and price above marginal cost. Such a result, which still holds with an arbitrary number of competing firms because a higher number of congregated firms makes the corresponding place even more attractive, may be viewed as the strategic counterpart of the main property of the CP model where firms also sell differentiated varieties. Thus, while the homogeneity of goods leads to dispersion of sellers, their heterogeneity favors their geographical concentration: Homogeneity leads to segmentation of markets because the firms can take advantage of their geographical isolation, while heterogeneity facilitates the overlapping of markets because consumers pay less attention to travel costs and more to specific attributes of sellers. In this case, firms no longer fear the effects of price competition (the dispersion force is weakened by product differentiation), while they strive to be as close as possible to consumers who seek differentiation (the agglomeration force). Since consumers are spread all over Main Street, firms set up at the location with the highest market potential, here at the market center, and therefore minimize their geographical differentiation.

At first sight, this extreme concentration of shops may seem inefficient but this needs qualification. At the social optimum, prices are set equal to marginal cost so that consumers’ well-being depends only upon firms’ locations. In the case of a homogeneous good, maximizing total welfare boils down to minimizing aggregate transportation costs. However, once differentiation is introduced across varieties, consumers no longer patronize the nearest firm on each trip because they value differentiation. Therefore, one needs a more general welfare function accounting for both the distance and product diversity effects (i.e., the consumers’ indirect utility). In this case, the formation of a cluster need not be socially suboptimal. Quite the opposite: When products are sufficiently differentiated, transport costs are low, or both, it is socially efficient to have a centrally located shopping district.

Concluding Remarks

At first sight, the contributions discussed in this article may seem like a patchwork of unrelated models. This impression may stem from the different modeling strategies used in spatial economics. But the fundamental question remains the same at all spatial scales:

Where do agents locate and why do they form spatial clusters? However, as many other questions are scale-specific (e.g., interregional interaction vs. intra-urban interaction), one needs different modeling strategies. Yet this cornucopia of models should not blind the reader to the existence of a few general ideas that build strong ties among different models.

Admittedly, shipping goods, migrating or commuting, and transmitting information involve very different costs. Internal and external increasing returns differ a great deal because the latter covers a wide array of phenomena that make the whole bigger than the sum of its parts. However, the trade-off between increasing returns and transport costs (broadly defined) is key to the understanding of the space-economy arising at different spatial scales, very much as different types of clusters are the outcome of the interplay between agglomeration and dispersion forces. Likewise, imperfect competition or perfect competition married with externalities both imply that market prices do not reflect the true social value of goods and services. Hence, a common feature of agglomeration models is the existence of a gap between market prices and social marginal costs. Therefore, the differences among models are not as marked as they may seem a priori.

In addition, spatial models share some common methodological features: They are seldom full-fledged general equilibrium models à la Arrow-Debreu or Marshallian partial equilibrium models of a specific market. Spatial models typically combine a few markets for particular commodities, such as land, labor, and transport, in order to study their interaction while the rest of the economy provides the numéraire under constant returns and perfect competition.

Despite many valuable contributions made since the 1990s, much more work remains to be done. By focusing on exchanges between regions to explain why some regions fare better than others, economic geography remains in the tradition of trade theory. By contrast, urban economics emphasizes the internal functioning of a city and does not address intercity trade much. Although both approaches are legitimate, full-fledged models of the regional-urban system should consider both. Therefore, the theory of urban systems seems to be the natural framework for the development of a synthesis.

Importantly, cities are not like Russian *matriochkas*. The nature, not just the size, of big cities differs from that of small or medium-sized cities. There is growing evidence that large cities are more skill-abundant than small cities (Behrens et al., 2014; Davis & Dinkel, 2019). Therefore, it is reasonable to expect certain effects to be at work in big cities but not in the others. What is more, cities specialize in different activities, have different sizes, and trade different goods. The pioneering work of Henderson (1974), which provides an elegant description on how cities of different sizes that trade different goods may emerge, has served as a foundation for a large body of research. Recent contributions, such as the work of Akamatsu et al. (2012) and spatial quantitative economics (Redding & Rossi-Hansberg, 2017), also open new avenues for research.

Acknowledgments

The author is very grateful to Masahisa Fujita and Stef Proost for the many discussions about the ideas and models surveyed in this article.

Further Reading

Ahlfeldt, G., Redding, S. J., Sturm, D., & Wolf, K. (2015). The economics of density: Evidence from the Berlin Wall. *Econometrica*, 83(6), 2127–2189.

Allen, T., & Arkolakis, C. (2014). Trade and the topography of the spatial economy. *Quarterly Journal of Economics*, 129(3), 1085–1040.

Bairoch, P. (1988). *Cities and economic development: From the dawn of history to the present*. Chicago, IL: University of Chicago Press.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G., & Robert-Nicoud, F. (2003). *Economic geography and public policy*. Princeton, NJ: Princeton University Press.

Cavallière, J., Gaigné, C., Tabuchi, T., & Thisse J. F. (2007). Trade and the structure of cities. *Journal of Urban Economics*, 62(3), 383–404.

Chauvin, J. P., Glaeser, E., Ma, Y., & Tobio, K. (2017). What is different about urbanization in rich and poor countries? Cities in Brazil, China, India and the United States. *Journal of Urban Economics*, 98, 17–49.

Combes, P.-P., Duranton, G., & Gobillon, L. (2011). The identification of agglomeration economies. *Journal of Economic Geography*, 11(2), 253–266.

Davis, D. R., & Weinstein, D. E. (2002). Bones, bombs, and break points: The geography of economic activity. *American Economic Review*, 92(5), 1269–89.

Duranton, G., Henderson, J. V., & Strange, W. C. (Eds.). (2015). *Handbook of regional and urban economics* (Vol. 5). Amsterdam, The Netherlands: Elsevier.

Fujita, M., & Thisse, J.-F. (2013). *Economics of agglomeration: Cities, industrial location and globalization* (2nd ed.). Cambridge, U.K.: Cambridge University Press.

Glaeser, E. L., & Gottlieb, J. D. (2009). The wealth of cities: Agglomeration economies and spatial equilibrium in the United States. *Journal of Economic Literature*, 47(4), 983–1028.

Glaeser, E. L., & Kahn, M. E. (2004). Sprawl and urban growth. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2481–2527). Amsterdam, The Netherlands: Elsevier.

Head, K., & Mayer, T. (2004). The empirics of agglomeration and trade. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2609–2669). Amsterdam, The Netherlands: Elsevier.

Henderson, J. V. (1988). *Urban development: Theory, fact and illusion*. Oxford, U.K.: Oxford University Press.

Henderson, J. V., Squires, T., Storeygard, A., & Weil, D. (2018). The global distribution of economic activity: Nature, history, and the role of trade. *Quarterly Journal of Economics*, 133(1), 357–406.

Ioannides, Y. M. (2012). *From neighborhoods to nations: The economics of social interactions*. Princeton, NJ: Princeton University Press.

Leamer, E. E. (2007). A flat world, a level playing field, a small world after all, or none of the above? A review of Thomas L. Friedman's *The World Is Flat*." *Journal of Economic Literature*, 45(1), 83–126.

Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1), 3–42.

Lucas, R. E., & Rossi-Hansberg, E. (2002). On the internal structure of cities. *Econometrica*, 70(4), 1445–1476.

Moretti, E. (2011). Local labor markets. In O. Ashenfelter & D. Card (Eds.), *Handbook of labor economics* (Vol. 4B, pp. 1237–1313). Amsterdam, The Netherlands: Elsevier.

Ottaviano, G. I. P., & Thisse, J.-F. (2004). Agglomeration and economic geography. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2563–2608). Amsterdam, The Netherlands: Elsevier.

Puga, D. (1999). The rise and fall of regional inequalities. *European Economic Review*, 43(2), 303–334.

Roback, J. (1982). Wages, rents, and the quality of life. *Journal of Political Economy*, 90(6), 1257–1278.

Robert-Nicoud, F. (2005). The structure of simple “new economic geography” models (or, on identical twins). *Journal of Economic Geography*, 5(2), 201–234.

Schweizer, U., & Varaiya, P. V. (1976). The spatial structure of production with a leontief technology. *Regional Science and Urban Economics*, 6(3), 231–251.

Zenou, Y. (2009). *Urban labor economics*. Cambridge, U.K.: Cambridge University Press.

References

Akamatsu, T., Takayama Y., & Ikeda, K. (2012). Spatial discounting, Fourier, and racetrack economy: A recipe for the analysis of spatial agglomeration models. *Journal of Economic Dynamics & Control*, 36(11), 1729–1759.

Alonso, W. (1964). *Location and land use*. Cambridge, MA: Harvard University Press.

- Bacolod, M., Bull, B. S., & Strange, W. C. (2009). Skills in the city. *Journal of Urban Economics*, 65(2), 136–153.
- Baldwin, R. E. (2016). *The great convergence*. Cambridge, MA: Harvard University Press.
- Behrens, K., Duranton, G., & Robert-Nicoud, F. (2014). Productive cities: Sorting, selection and agglomeration. *Journal of Political Economy*, 122(3), 507–553.
- Behrens, K., & Robert-Nicoud, F. (2015). Agglomeration theory with heterogeneous agents. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 171–245). Amsterdam, The Netherlands: Elsevier.
- Charlot, S., Gaigné, C., Robert-Nicoud, F., & Thisse, J.-F. (2006). Agglomeration and welfare: The core-periphery model in the light of Bentham, Kaldor, and Rawls. *Journal of Public Economics*, 90(1–2), 325–347.
- Chetty, R., & Hendren, N. (2018). The effects of neighborhoods on intergenerational mobility I: Childhood exposure effects. *Quarterly Journal of Economics*, 133(3), 1107–1162.
- Combes, P.-P., Duranton, G., & Gobillon, L. (2008). Spatial wage disparities: Sorting matters! *Journal of Urban Economics*, 63(2), 723–742.
- Combes, P.-P., Duranton, G., Gobillon, L., Puga, D., & Roux, S. (2012). The productivity advantages of large cities: Distinguishing agglomeration from firm selection. *Econometrica*, 80(6), 2543–2594.
- Combes, P.-P., & Gobillon, L. (2015). The empirics of agglomeration economies. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 247–348). Amsterdam, The Netherlands: Elsevier.
- d’Aspremont, C., Gabszewicz, J. J., & Thisse, J.-F. (1979). On Hotelling’s “stability in competition.” *Econometrica*, 47(5), 1045–1050.
- Davis, D. R., & Dingel, J. (2019). A spatial knowledge economy. *American Economic Review*, 109(1), 153–170.
- de Palma, A., Ginsburgh, V., Papageorgiou, Y. Y., & Thisse, J.-F. (1985). The principle of minimum differentiation holds under sufficient heterogeneity. *Econometrica*, 53(4), 767–781.
- Duranton, G., & Puga, D. (2004). Micro-foundations of urban increasing returns: Theory. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics: Cities and geography* (pp. 2063–2117). Amsterdam, The Netherlands: Elsevier.
- Duranton, G., & Puga, D. (2015). Urban land use. In G. Duranton, J. V. Henderson, & W. C. Strange (Eds.), *Handbook of regional and urban economics* (Vol. 5, pp. 467–560). Amsterdam, The Netherlands: Elsevier.

Faggio, G., Silva, O., & Strange, W. C. (2017). Heterogeneous agglomeration. *Review of economics and statistics*, 99(1), 80–94.

Fujita, M. (1989). *Urban economic theory: Land use and city size*. Cambridge, U.K.: Cambridge University Press.

Fujita, M., Krugman, P., & Venables, A. J. (1999). *The spatial economy: Cities, regions and international trade*. Cambridge, MA: MIT Press.

Fujita, M., & Thisse, J.-F. (2003). Does geographical agglomeration foster economic growth? And who gains and loses from it? *Japanese Economic Review*, 54(2), 121–145.

Fujita, M., & Thisse, J.-F. (2006). Globalization and the evolution of the supply chain: Who gains and who loses? *International Economic Review*, 47(3), 811–836.

Gibbons, S., Overman, H. G., & Pelkonen, P. (2014). Area disparities in Britain: Understanding the contribution of people vs. place through variance decompositions. *Oxford Bulletin of Economics and Statistics*, 76(3), 745–763.

Glaeser, E. L., & Gyourko, J. (2018). The economic implications of housing supply. *Journal of Economic Perspectives*, 32(1), 3–30.

Glaeser, E. L., Kolko, J., & Saiz, A. (2001). Consumer city. *Journal of Economic Geography*, 1(1), 27–50.

Glaeser, E. L., Renssler, M. G., & Tobio, K. (2008). *Urban inequality* (NBER Working Paper 14419).

Handbury, J., & Weinstein, D. E. (2015). Goods prices and availability in cities. *Review of Economic Studies*, 82(1), 258–296.

Head, K., & Mayer, T. (2011). Gravity, market potential and economic development. *Journal of Economic Geography*, 11(2), 281–294.

Head, K., & Mayer, T. (2014). Gravity equations: Workhorse, toolkit, and cookbook. In G. Gopinath, E. Helpman, & K. Rogoff (Eds.), *Handbook of international economics* (Vol. 4, pp. 131–196). Amsterdam, The Netherlands: Elsevier.

Helpman, E. (1998). The size of regions. In D. Pines, E. Sadka, & I. Zilcha (Eds.), *Topics in public economics: Theoretical and applied analysis* (pp. 33–54). Cambridge, U.K.: Cambridge University Press.

Henderson, J. V. (1974). The sizes and types of cities. *American Economic Review*, 64(4), 640–656.

Henderson, J. V. (1997). Medium size cities. *Regional Science and Urban Economics*, 27(6), 583–612.

- Henderson, J. V., & Ono, Y. (2008). Where do manufacturing firms locate their headquarters? *Journal of Urban Economics*, 63(2), 431–450.
- Hotelling, H. (1929). Stability in competition. *Economic Journal*, 39(153), 41–57.
- Inglehart, R. F., & Norris, P. (2016). *Trump, Brexit, and the rise of populism: Economic have-nots and cultural backlash* (Harvard Kennedy School WP 16-026).
- Kaldor, N. (1935). Market imperfection and excess capacity. *Economica*, 2(5), 35–50.
- Krugman, P. (1980). Scale economies, product differentiation, and the pattern of trade. *American Economic Review*, 70(5), 950–959.
- Krugman, P. (1991). Increasing returns and economic geography. *Journal of Political Economy*, 99(3), 483–499.
- Krugman, P., & Venables, A. J. (1995). Globalization and the inequality of nations. *Quarterly Journal of Economics*, 110(4), 857–880.
- Manski, C. (2000). Economic analysis of social interactions. *Journal of Economic Perspectives*, 14(3), 115–136.
- McFadden, D. (1986). The choice theory approach to market research. *Marketing Science*, 5(4), 275–297.
- Mills, E. S. (1967). An aggregative model of resource allocation in a metropolitan area. *American Economic Review*, 57(2), 197–210.
- Miyagiwa, K. (1991). Scale economies in education and the brain drain problem. *International Economic Review*, 32(3), 743–759.
- Moretti, E. (2004). Human capital externalities in cities. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2243–2291). Amsterdam, The Netherlands: Elsevier.
- Moretti, E. (2012). *The new geography of jobs*. Boston, MA: Houghton Mifflin Harcourt.
- Murata, Y. (2003). Product diversity, taste heterogeneity, and geographic distribution of economic activities: Market vs. non-market interactions. *Journal of Urban Economics*, 53(1), 126–144.
- Murata, Y., & Thisse, J.-F. (2005). A simple model of economic geography à la Helpman-Tabuchi. *Journal of Urban Economics*, 58(1), 137–155.
- Ogawa, H., & Fujita, M. (1980). Equilibrium land use patterns in a non-monocentric city. *Journal of Regional Science*, 20(4), 455–475.

Redding, S. (2011). Economic geography: A review of the theoretical and empirical literature. In D. Bernhofen, R. Falvey, D. Greenaway, & U. Kreickemeie (Eds.), *The Palgrave handbook of international trade* (pp. 497–531). London, U.K.: Palgrave Macmillan.

Redding, S. J., & Rossi-Hansberg, E. (2017). Quantitative spatial economics. *Annual Review of Economics*, 9, 21–58.

Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2119–2171). Amsterdam, The Netherlands: Elsevier.

Starrett, D. (1978). Market allocations of location choice in a model with free mobility. *Journal of Economic Theory*, 17(1), 21–37.

Tabuchi, T., Thisse, J.-F., & Zhu, X. (2018). Does technological progress magnify regional disparities? *International Economic Review*, 59(2), 647–663.

Takahashi, T., Takatsuka, H., & Zeng, D.-Z. (2013). Spatial inequality, globalization, and footloose capital. *Economic Theory*, 53(1), 213–238.

von Thünen, J. H. (1826/1966). *Der Isolierte Staat in Beziehung auf Landwirtschaft und Nationalökonomie* [The Isolated State]. Oxford, U.K.: Pergamon Press.

Jacques-François Thisse

CORE, Université Catholique de Louvain