DISCUSSION PAPER SERIES

DP13638

NEW ECONOMIC GEOGRAPHY AND THE CITY

Carl Gaigné and Jacques-François Thisse

INTERNATIONAL TRADE AND REGIONAL ECONOMICS

ISSN 0265-8003

NEW ECONOMIC GEOGRAPHY AND THE CITY

Carl Gaigné and Jacques-François Thisse

Discussion Paper DP13638
Published 01 April 2019
Submitted 27 March 2019

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

NEW ECONOMIC GEOGRAPHY AND THE CITY

Abstract

In this chapter, we provide a bird-eye overview of recent developments in NEG within a unifying framework. We build on the idea that the difference in the economic performance of regions depend on the global and local interactions between and within regions through the locational decisions made by firms and households at the macro and microspatial levels. We also focus on settings that take into account the urban structure, the social and skill composition and the sectorial specialization of regional agglomerations, and the quality of urban life. Three types of spatial frictions are considered, that is, transport costs, commuting costs, and communication costs.

JEL Classification: L12, O14, R12

Keywords: new economic geography, cities, transport costs, Commuting costs, Communication costs, Land rent, manufacturing goods, Services

Carl Gaigné - carl.gaigne@inra.fr
INRA, UMR1302, SMART-LERECO Rennes (France), and CREATE, Laval University, Quebec (Canada).

Jacques-François Thisse - jacques.thisse@uclouvain.be
CORE-UCLouvain (Belgium), Higher School of Economics (St Petersburg) and CEPR

Acknowledgements
We thank P. Asadi, M. Fujita, T. Gokan, S. Kichko, F. Moizeau, D. Peeters, S. Riou, and T. Tabuchi for their comments and suggestions. This work is supported by the Russian Science Foundation (Project No. 18-18-00253).

New Economic Geography and the City*

Carl Gaigné† Jacques-François Thisse‡

March 27, 2019

Abstract

In this chapter, we provide a bird-eye overview of recent developments in NEG within a unifying framework. We build on the idea that the difference in the economic performance of regions depend on the global and local interactions between and within regions through the locational decisions made by firms and households at the macro and microspatial levels. We also focus on settings that take into account the urban structure, the social and skill composition and the sectorial specialization of regional agglomerations, and the quality of urban life. Three types of spatial frictions are considered, that is, transport costs, commuting costs, and communication costs.

Keywords: new economic geography, cities, transport costs, commuting costs, communication costs, land rent, manufacturing goods, services

JEL Classification: L12, O14, R12

*We thank P. Asadi, M. Fujita, T. Gokan, S. Kichko, F. Moizeau, D. Peeters, S. Riou, and T. Tabuchi for their comments and suggestions. This work is supported by the Russian Science Foundation (Project No. 18-18-00253).

†INRA, UMR1302, SMART-LERECO Rennes (France), and CREATE, Laval University, Quebec (Canada).

‡CORE-UCLouvain (Belgium), National Research University Higher School of Economics (Russian Federation), and CEPR.

1 Introduction

Ever since the publication of Krugman's (1991) pioneering paper, New Economic Geography (NEG) has given new life to spatial economics, which since then has made enormous progress by any previous yardstick. The very name "New Economic Geography" seems chosen to stir a debate: Is NEG economic geography proper or rather spatial economics? And is there anything really new in it? To the best of our knowledge, no economist before Krugman had been able to show how regional imbalances can arise within the realm of general equilibrium theory. To achieve this, Krugman has borrowed concepts and tools developed in modern economic theory, especially the Dixit and Stiglitz (1977) model of monopolistic competition, which is the workhorse of new growth and trade theories. As for transport costs, Krugman uses the iceberg technology: Only a fraction of a good shipped between two places reaches the destination, the missing share having melted on the way. This ingenious modeling trick, due to Samuelson (1954), allows integrating positive shipping costs without having to deal explicitly with a transport sector. Hence, Dixit, Stiglitz and Samuelson form the trinity under which Krugman has combined increasing returns, commodity trade and the mobility of production factors within his now famous "core-periphery" model.

In NEG, the distribution of activities emerges as the unintentional outcome of a myriad of decisions made by firms and households pursuing their own interest. Thus, *methodologically NEG belongs to mainstream economics*. This is probably what distinguishes most NEG from economic geography proper. Our choice to focus on NEG only does not reflect any prejudice on our part. It is mainly driven by the need to stress how this approach can be used to highlight old and new issues. Being deeply rooted in mainstream economics, NEG has strong connections with various economic fields, including industrial organization and international trade, but also with the new theories of growth and development. This permits cross-fertilizations that have been out of reach for a long time. Furthermore, in terms of its subject matter, NEG cannot be considered alien to regional science as many ideas and concepts NEG builds on have been around for a long time (Ottaviano and Thisse 2005). For example, the fundamental idea that the interplay between different types of scale economies and transport costs is critical for the way the space-economy is organized, and so at various spatial scales (cities, regions, countries, continents), was known (at least) since the work of Weber and Lösch. It is fair to say, however, that those ideas were fairly disparate and in search of a synthesis.

It is widely recognized that Krugman's main contribution was his entirely different and new approach to regional disparities. Under constant returns, firms find it profitable to disperse their production to bring it closer to customers, as this will reduce transport costs without lowering productive efficiency. Such a space-economy is the quintessence of self-sufficiency: If the distribution of factor endowments is uniform, the economy reduces to a Robinson Crusoe-type economy where each person produces for her own consumption. Under these circumstances, only differences in endowments of immobile production factors can explain the marked differences in the spatial distribution of activities, and hence the need for interregional and international trade.

Relying only on first nature to explain the existence of large urban agglomerations and sizable trade flows amounts to playing Hamlet without the Prince. Krugman squarely tackles this problem by assuming that firms operate under increasing returns and imperfect competition on the product market. Such a combination is orthogonal to the standard paradigm with constant returns and perfect competition, which has dominated economic theory for a long time. Furthermore, to the trade-off between increasing returns and transport costs Krugman (1980) has added a third variable: The size of spatially separated markets. The main accomplishment of NEG has been to highlight *how market size interacts with scale economies and transport costs to shape the space-economy.*

In NEG, the market outcome stems from the interplay between a dispersion force and an agglomeration force operating within a full-fledged general equilibrium model. In Krugman (1991) and Fujita et al. (1999a), the dispersion force stems from the spatial immobility of farmers whose demands for the manufactured good are to be met. The agglomeration force is more involved and requires a more detailed description. If a larger number of manufactures is located in one region, the number of varieties locally produced is larger too. Then, manufactured goods are available at lower prices in the larger region because local varieties are cheaper than imported varieties. This in turn induces consumers living in the smaller region to move toward the larger region, where they may enjoy a higher standard of living. The resulting increase in the numbers of consumers creates a larger demand for the manufactured good, which, therefore, leads additional firms to locate therein. This implies the availability of more varieties in the region in question but less in the other because there are scale economies at the firm's level. Consequently, as noticed by Krugman (1991, p.486), *circular causation* is present because these two effects reinforce each other: "manufactures production will tend to concentrate where there is a large market, but the market will be large where manufactures production are concentrated." The great accomplishment of Krugman was to integrate all these various effects within a unified framework and to show that the level of *transport costs* is the key-determining factor for the organization of the space-economy.

When transport costs are sufficiently low, Krugman (1991) showed that manufactures are concentrated in a single region that becomes the *core* of the economy, whereas the other region, called the *periphery*, supplies only the agricultural good. Firms are able to exploit scale economies by selling more in the larger market without losing much business in the smaller market because transport costs are low. For exactly the opposite reasons, the economy displays a symmetric regional pattern of production when transport costs are high: Because the local markets are now protected by geographical separation, firms relax competition by being dispersed across regions. Hence, the core-periphery model 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. It is also worth stressing that the dual structure made of a core and a periphery is brought about by market forces acting in a set up formed by two regions that are ex ante identical. Note, finally, that the above results hold true in more general settings. For example, Fujita and Krugman (1994) and Fujita et al. (1999b) have extended Krugman's set up to describe the formation of urban systems in an economy where agriculture is the dominant sector. The reader is referred to Fujita et al. (1999a), Baldwin et al. (2003), and Fujita and Thisse (2013) for detailed analyses of settings that will not be discussed here.

By focusing on the interaction between the product and labor markets, Krugman's work remains in the tradition of international trade. Although we recognize the limits of this approach, we believe that it delivers a powerful framework that has proven to be very useful in dealing with a large number of issues. However, the core-periphery model seems to be of less relevance to describe on-going changes in developed economies because it focuses more on the trade causes of the agglomeration of activities, and not much on its urban sources. Yet, housing and commuting costs, which are the two quintessentially urban commodities, account for a large and growing share of consumers' expenditures. For example, in the United States housing accounts on average for 20% of household budgets while 18% of total expenditures is spent on car purchases, gasoline, and other related expenses that do not include the cost of time spent in traveling. According to the Census, 139 million American workers have spent a collective 3.4 million years in commuting during the year 2014. Albouy et al. (2018) show that, in 2006, the total value of urban land in the U.S. exceeds twice the American GDP, while in most countries the value of land keeps rising (Knoll et al. 2017). All of this leads us to concur with Helpman (1998) and Tabuchi (1998) and to assert that *housing and commuting costs are the main dispersion force at work in modern economies.*

In this chapter, we do not discuss the canonical NEG models because they ignore intra-regional interactions. Rather, we build on the idea that the difference in the economic performance of regions is to a large extent explained by the behavior and interactions between households and firms located in cities. The focus thus shifts from the *nation-state* to the *city-region*. In other words, we discuss NEG-like models in which there is a land market, but no agricultural sector. The common thread of this chapter may thus be viewed as an attempt to unify Alonso and Krugman. Consequently, we need a framework that differs from Krugman's. This alternative framework should also display enough versatility to revisit new issues such as the quality of urban life, the decentralization of jobs, spatial segregation, and spillovers through the interaction between three types of spatial frictions, i.e., transport costs, commuting costs, and communication costs. This is achieved here by using the setting developed by Tabuchi and Thisse (2006) where regions have a geographical dimension.

Competition for land among consumers gives rise to land rent and commuting costs that both increase with population size. As a result, the space-economy is the outcome of the interaction between two types of spatial costs: The transport costs of commodities *between* cities and the commuting costs borne by workers *within* cities. The results presented in Section 2 differ from those obtained by Krugman: The evolution of commuting costs within cities, instead of transport costs between cities, becomes the key-factor explaining how the space-economy is organized. Moreover, despite the many advantages provided by the inner city through an easy access to highly specialized services, the abyssal fall in communication costs has led firms or developers to form enterprise zones or edge cities (Henderson and Mitra 1996). We then go one step further by allowing firms to form secondary business centers. The analysis shows how a polycentric city may be able to reduce urban costs, which in turn permits a large city to retain its dominant position. In the same vein, we will also show that manufactures may choose to keep their strategic functions in large metropolitan areas while relocating their production plants in remote areas where land and labor are cheap.

The bulk of the NEG literature has concentrated on manufacturing sectors, although employment in modern cities is found mainly in firms providing untradable consumption and business services. While the Industrial Revolution fostered the emergence of manufacturing cities, services show a taste for cities that manufacturing sectors no longer have. As stressed by Glaeser et al. (2001), the success of a city depends more than before on its role as a center of consumption, that is, on the supply of local amenities and services. In Section 3, we recognize that most consumption services (e.g., health care, restaurants and movie theaters) are untradable and study their intercity distribution. In Section 4, we blend a service sector and a manufacturing sector in an economy in which workers display both *sectorial and spatial mobility.*

We want to stress that the ideas discussed above are presented through the lenses of a new framework. Instead of using the CES model of monopolistic competition, we consider a linear model of monopolistic competition that has been developed in industrial organization (Vives 1999). This setting allows us to build NEG models that are analytically solvable by means of paper and pencil, something hard to achieve with the CES. Our framework also allows us to discuss new issues, which are especially hard to handle within the standard NEG models. As for the transport cost, it is added to the production cost and measured in terms of the numéraire. This modeling strategy avoids imposing binding relationships between prices and shipping costs. By yielding linear equilibrium conditions, this model delivers a full analytical solution that captures in a simple way the pro-competitive effects associated with market size and market integration.

To accomplish this task, we use a NEG-like model that takes into account the following fundamental aspects of urban development: (i) Cities can be monocentric or polycentric; (ii) cities are anchored, separated by a given physical distance, and connected through their main centers, (iii) cities supply housing and untradable services, as well as tradable goods; and (iv) cities host workers and retirees who are attracted by different location factors. We are well aware that the reader accustomed to NEG might be surprised by our choice of menu. It is worth stressing that the model used in this chapter can replicate the main results obtained by Krugman and others (Ottaviano and Thisse 2004). It thus belongs to NEG. The seemingly different approach followed here has been chosen in the hope of convincing the skeptical regional scientist and geographer that NEG is a lively field that still has a high potential for future research.

The lack of attention paid by economists to earlier contributions in spatial analysis is unwarranted. Regional scientists and geographers have developed several models, such as those ranging from the entropy to the gravity and logit models, which have proven to be very effective in predicting and explaining different types of flows (Wilson 2010). By ignoring this body of research, economists have sometimes rediscovered the wheel and missed the opportunity of developing earlier a sound theory of the space-economy. But equally unwarranted is the acrimony expressed by many geographers soon after the diffusion of Krugman's work: They miss the importance of working with a fully consistent microeconomic model, especially the need of using a well-defined market structure and a precise specification of the agglomeration effects at work.

The remainder of this chapter reflects the above methodological choices. The baseline model is presented in Section 2 and the market outcome is compared to Krugman's core-periphery structure. We show how commuting costs and population density affect the location of economic activities *between* and *within* cities. Sections 3 and 4 focus, respectively, on untradable services and on their interactions with tradable manufactured goods. To illustrate the potential of NEG in the study of policy-driven problems, Section 5 addresses the relationship between cities and crime, the environmental and economic consequences of compact cities, and the implications of an aging population for the urban system. The move towards globalization is the outcome of a process in which various types of mobility costs have substantially decreased. So, we will also briefly discuss the impact of falling communication costs. Section 6 concludes.

2 Cities and Tradable Goods

The economy is formed by two cities/regions, labelled $r = A, B$, and populated by L consumers who are free to choose where to live. Unlike the standard core-periphery model where regions are spaceless places, we recognize explicitly that any sizeable human settlement takes the form of a city where economic agents work and reside.

There is one sector and three goods: Land (housing), a manufactured good, which is differentiated and tradable, and an unproduced homogeneous good, which is the numéraire. Space is one-dimensional and the opportunity cost of land is zero. Each region can be urbanized by accommodating firms and consumers according to rules described below. Whenever a city exists, it has a *central business district* (CBD) where firms set up. Since NEG has nothing really new to add to the reasons explaining why CBDs exist, it is convenient to assume that CBDs pre-exist and have no spatial extension. There are also transport and commuting costs.

The main reason for the existence of cities is the presence of increasing returns. Under scale economies internal to firms, consumers have direct access to the locally produced varieties, the number of which depends on the city size. When they display a love for variety, consumers are also inclined to consume varieties produced in other places. This in turn prompts trade in differentiated commodities across spatially separated markets. However, foreign varieties must be imported at a positive transport cost, which tends to make them more expensive.

The standard thought experiment of NEG is well known: How do firms and consumers locate when the cost of shipping the manufactured good *between* regions/cities steadily decreases? Once we account for a description of the space-economy that fits better the contemporary world than the standard setting of NEG, this thought experiment must be supplemented by another one, that is, the impact of commuting costs *within* cities. In sum, both the transport costs of commodities and the commuting costs of people must be taken into account to understand how economic activities are distributed across space.

2.1 The Spatial Economy

We assume that the lot size is fixed and normalized to 1. This assumption does not allow replicating the well-documented fact that the population density is higher in the city center than in the suburbs. It is widely used, however, in models involving an urban economics building block because it captures the basic trade-off between long/short commutes and low/high land rents. Furthermore, the housing floor space per unit of land of buildings is given by δ, which is the same across the city. The parameter δ measures the city's compactness through the population density.

Let λ_r be the share of consumers living in city r. At the residential equilibrium all consumers reach the same utility level. If land is available on both sides of the CBD, the residential equilibrium involves a symmetric distribution of consumers around the CBD with city r's right hand side limit at

$$\bar{x}_r = \frac{\lambda_r L}{2\delta}.$$

In other words, the city is described by an interval of length $\lambda_r L/\delta$. In this chapter, we treat the tallness of buildings parametrically. However, in a more general setting, δ should be determined endogenously through a given production function (Fujita 1989).

A. Consumers. Consumers share the same quasi-linear preferences. Denoting by n the total mass of varieties, the utility derived from consuming q_i units of variety $i \in [0, n]$ is given by

$$u(q_i) = \alpha q_i - \frac{\beta}{2} q_i^2 - \frac{\gamma}{2n} q_i \int_0^n q_j \mathrm{d}j. \tag{1}$$

The parameters α, β, and γ are interpreted as follows: $\alpha > 0$ measures the desirability of the manufactured good with respect to the numéraire, whereas γ $(0 < \gamma < \beta)$ is a measure of the degree of substitution between varieties. Moreover, (1) shows that the marginal utility of variety i decreases with its own consumption as well as with the total consumption of the manufactured good.

Preferences over all varieties are obtained by nesting the subutility (1) into a linear utility:

$$U(q_0, q_{i\le n}) = \alpha \int_0^n q_i \mathrm{d}i - \frac{\beta}{2} \int_0^n q_i^2 \mathrm{d}i - \frac{\gamma}{2n} \int_0^n q_i \left(\int_0^n q_j \mathrm{d}j \right) \mathrm{d}i + q_0, \tag{2}$$

where q_0 is the quantity of the numéraire. The second term in (2) has the nature of a Herfindahl index, thus meaning that $\beta > 0$ amounts to assuming that the consumer have a love for variety. Indeed, when an individual consumes nq units of the manufactured such that the consumption is uniform and equal to nq/x on $[0, x]$ and zero on $(x, n]$, she enjoys the utility level

$$\begin{aligned} U &= \alpha \int_0^x \frac{nq}{x} \mathrm{d}i - \frac{\beta}{2} \int_0^x \left(\frac{nq}{x}\right)^2 \mathrm{d}i - \frac{\gamma}{2n} \int_0^x \int_0^x \left(\frac{nq}{x}\right)^2 \mathrm{d}i\mathrm{d}j + q_0 \\ &= \alpha nq - \frac{\beta}{2x} n^2 q^2 - \frac{\gamma}{4n} n^2 q^2 + q_0. \end{aligned}$$

This function is increasing in x and, hence, maximized at $x = n$ where variety consumption is maximal. Consumers endowed with preferences render large markets more attractive as long as locally produced varieties are cheaper than imported varieties.

The lot size being fixed, there is no need to specify how housing space enters into preferences. In what follows, we ease the burden of notation by adopting two normalizations that entail no loss of generality: The unit of the numéraire is chosen for $\alpha = 1$ and the unit of the manufactured good for $\gamma = 1$ to hold.

Each consumer supplies inelastically one unit of labor. Consumers commute to the CBD where jobs are located and earn an income w_r, which is determined at the equilibrium. The unit commuting cost is given by $\theta > 0$, and thus a consumer located at $x > 0$ bears a commuting cost equal to θx units of the numéraire. In addition, each consumer is endowed with $\bar{q}_0 > 0$ units of the numéraire, which is sufficiently large for the individual consumption of the homogeneous good to be strictly positive at the equilibrium outcome.

It remains to close the model by specifying the structure of land ownership. Since we use quasi-linear preferences, the demand for the manufactured good does not depend on consumers' incomes. Therefore, the way the land rent is redistributed affects only the consumption of the numéraire. In what follows, we follow the main approach used in urban economics and assume than land is collectively owned, so that the aggregate land rent is evenly distributed across consumers.

The budget constraint faced by a consumer living in city r is given by

$$\int_0^n q_{ir} p_{ir} \mathrm{d}i + q_0 + R_r(x)/\delta + \theta x = w_r + \bar{q}_0, \tag{3}$$

where p_{ir} (q_{ir}) is the price (consumption) of variety i in city r, $R_r(x)$ the land rent at x so that $R_r(x)/\delta$ is the price paid by a consumer to reside at x.

A consumer chooses her location and consumption bundle so as to maximize her utility (2) subject to the budget constraint (3). This yields the following individual demand for variety i in city r (Tabuchi and Thisse 2006):

$$q_{ir} = \frac{1}{\beta + 1} - \frac{p_{ir}}{\beta} + \frac{1}{\beta(\beta + 1)} \bar{p}_r \qquad \text{with} \quad \bar{p}_r \equiv \frac{1}{n} \int_0^n p_{kr} \mathrm{d}k, \tag{4}$$

where $\bar{p}_r$ is the average price of all varieties in city r. The demand (4) captures in the very simple way the impact of competition on a firm's demand: A higher (lower) average price shifts upward (downward) the demand for variety i because local competition is softer (tougher), thus making variety i more (less) attractive to city r-consumers. Furthermore, because each firm is negligible, the price p_{ir} chosen by firm i has no impact on the average price $\bar{p}_r$.

The gains from using (2) are reaped in the various issues that can be addressed by using our baseline model. These issues, which are hard to handle under CES preferences, are discussed below.

B. Land market and urban costs. We call *urban costs* the sum of housing and commuting costs borne by a city r-consumer residing at any location x:

$$UC_r(x) = R_r(x)/\delta + \theta x.$$

Let $\Psi_r(x)$ be the highest price a worker is willing to pay to reside at location x in city r. Since there is only one type of labor, the equilibrium land rent is equal to $R_r^*(x) = \delta \max\{\Psi_r(x), 0\}$. A longer commute is equal to the decrease in the bid rent Ψ_r because $\partial\Psi_r/\partial x + \theta = 0$. Hence, $\Psi_r(x) = k - \theta x$ where k is a constant to be determined. Since the land rent at $\bar{x}_r$ is equal to the opportunity cost of land, here zero, we have $k = \theta\bar{x}_r$, and thus $\Psi_r(x) = \theta\lambda_r L/2\delta - \theta x$. Therefore, the price paid to reside at x is the mirror image of the corresponding commuting costs:

$$\frac{R_r^*(x)}{\delta} = \theta\left(\frac{\lambda_r L}{2\delta} - x\right) \qquad \text{for } x < \bar{x}_r.$$

Consequently, the price paid by a consumer to live at $x > 0$ decreases with the height of buildings because the average commuting is shorter.

The urban costs borne by consumers in city r do not depend their residential location within this city and are equal to

$$UC_r(\lambda_r) = \frac{\theta}{\delta}\frac{\lambda_r L}{2}. \tag{5}$$

Since they increase with city size, *urban costs act as the dispersion force.* As expected, intercity differences in urban costs increase with θ and decrease with δ because more consumers can share the cost of residing at a given distance x to the CBD.

C. Producers. Firms produce a differentiated and tradable good under monopolistic competition and increasing returns; for simplicity, they do not use land. A firm produces a single variety and any two firms supply two different varieties. Producing a variety of the manufactured good requires a zero marginal requirement and a positive fixed requirement ϕ of labor units. Hence, the total mass of varieties supplied in the economy is given by $n = L/\phi$ and the mass of firms producing in city r by $n_r = \lambda_r L/\phi$, so that λ_r is also the share of firms located in city r. The value of ϕ may be interpreted as an inverse measure of labor productivity.

Markets are segmented, that is, each firm is able to set a price specific to the market in which its output is sold. Since preferences and technologies are symmetric, firms sell their varieties at the same price in each city. Thus, we may disregard the index i and write the profits earned by a city r-firm as follows ($s \neq r$):

$$\pi_r = p_{rr} q_{rr}(p_{rr}) \lambda_r L + (p_{rs} - \tau) q_{rs}(p_{rs}) \lambda_s L - \phi w_r, \tag{6}$$

where p_{rr} is the price set by the local firms and p_{rs} the consumer price charged in city s by the firms located in $r \neq s$; q_{rr} and q_{rs} are the quantities sold by the r-firms in cities r and s, respectively.

The average price in city r is given by $\bar{p}_r = \lambda_r p_{rr} + \lambda_s p_{sr}$. Plugging this expression into (4) and solving the first-order conditions yield the equilibrium prices:

$$p^*_{rr} = \frac{2\beta + \tau(1 - \lambda_r)}{2(2\beta + 1)} \qquad p^*_{rs} = p^*_{ss} + \frac{\tau}{2}. \tag{7}$$

Observe that p^*_{rr} depends on λ_r and τ because the share of local firms matter while transport costs affect p^*_{rr} through the consumer price charged by the foreign firms. Both prices p^*_{rr} and p^*_{rs} capture the pro-competitive effects associated with a higher share of local competitors and lower transport costs. In particular, the prices of local varieties are lower in the larger city than in the smaller because competition is tougher. This result runs against the conventional wisdom that tradables are more expensive in larger cities because land rents and wages are higher therein. This argument overlooks the following two effects: competition is tougher in the larger city, while the quality improvements of existing goods and the introduction of new goods allow consumers to substitute low-priced goods for high-priced goods. Controlling for these effects, Handbury and Weinstein (2015) use a dataset covering $10 - 20$ million purchases of grocery items and find that prices for the same goods are equal or lower in larger cities. This highlights a trade-off that has been neglected in the urban economics literature: Consumers bear higher urban costs in larger cities, but tradable goods may be available at lower prices.

Furthermore, (7) also shows that lowering transport costs exacerbates competition in each city though the consumer price of imported varieties is higher than domestic varieties' because distant firms have to cover the cost of shipping their output. Therefore, consumption is biased toward locally produced goods. By contrast, the producer price of imported varieties is smaller than that of local varieties. The r-firms choose a pass-through smaller than 1 to facilitate the penetration of their varieties in city s. In other words, there is *spatial price discrimination.*

Finally, for intercity trade to occur and its pro-competitive effects to become concrete, transport costs cannot be too high: $p^*_{rs} - \tau > 0$. This condition holds regardless of the spatial distribution of firms if and only if $\tau < \tau_{\text{trade}} \equiv 2\beta/(2\beta + 1) < 1$.

D. Labor market and wages. Urban labor markets are local while labor market clearing implies that the creation and destruction of firms is governed by the location of consumers. Specifically, the equilibrium wage is determined by a bidding process in which potential firms compete for workers by offering them higher and higher wages until no firm can profitably enter the market. Put simply, operating profits are completely absorbed by the wage bill. The equilibrium quantities sold are given by $q^*_{rr} = p^*_{rr}/\beta$ and $q^*_{rs} = (p^*_{rs} - \tau)/\beta$. Plugging the equilibrium prices and quantities into π_r and solving for w_r give the equilibrium wage in city r, namely $w^*_r = \Pi_r/\phi$ where Π_r are the operating profits made in r.

The impact of λ_r on w^*_r is a priori unclear. Indeed, as λ_r rises, city r's size gets bigger, but the local firms' markup is smaller. By contrast, the size of city s gets smaller while the foreign firms' markup is higher. However, it is readily check that the equilibrium wage is higher when firms and workers are agglomerated in city r rather than equally dispersed between the two cities. This wage premium is caused here by market size, not by agglomeration economies. Note that the consumer surplus in A evaluated at the equilibrium is also higher under agglomeration than dispersion because all firms are located in A, while urban costs take their highest value under agglomeration.

2.2 The Formation of Manufacturing Cities

The size of the product and labor markets is endogenous when consumers are mobile. Indeed, when consumers move from one city to the other, they bring with them both their production and consumption capacities. As a consequence, the numbers of consumers and workers change in the two cities.

The locational choice made by a consumer is driven by the indirect utility associated with (2):

$$V_r(\lambda_r) = CS_r + w^*_r - UC_r + q^*_0, \tag{8}$$

where CS_r is the consumer surplus and q^*_0 the equilibrium consumption of the numéraire in city r. Hence, when choosing the city where she lives a consumer takes into account the income she earns, the level of urban costs she bears, and the consumer surplus she enjoys in the city though the supply of local and imported varieties. Thus, though the individual demands (4) are unaffected by income, the migration decision takes income into account. Everything else equal, workers are pulled by the higher wage region. The population becoming larger, the local demand for the manufactured good is raised, which attracts additional firms.

Although the present framework differs from Krugman's (1991), it captures the same effects. It also encapsulates the following fundamental trade-off, which is absent in Krugman: Concentrating people and firms in a small number of large cities minimizes the cost of shipping commodities among urban areas, but makes work-trips (as well as many other within-city trips) longer; when dispersion prevails, consumers bear lower commuting costs, but goods are more expensive because each city produces a smaller number of varieties.

At the equilibrium, each worker maximizes her utility subject to her budget constraint, each firm maximizes its profits, and markets clear. In particular, no consumer/firm has an incentive to change place. Denoting by $\lambda \equiv \lambda_A$ the endogenous population in city A, a *spatial equilibrium* arises at $1/2 \leq \lambda^* < 1$ when the utility differential $\Delta V(\lambda^*) \equiv V_A(\lambda^*) - V_B(\lambda^*)$ is equal to 0. When $\Delta V(1) \geq 0$, $\lambda^* = 1$, and thus all consumers and firms are set up in city A. NEG models typically display several spatial equilibria. In such a context, it is convenient to use stability as a selection device since an unstable equilibrium is unlikely to happen. An interior equilibrium is *stable* if, for any marginal deviation away from the equilibrium, the incentive system provided by the market brings the distribution of consumers back to the original one. This is so if and only if the slope of the utility differential ΔV is strictly negative at λ^*. By contrast, an agglomerated equilibrium is stable whenever it exists.

Replacing each term of V_r by its expression leads to the following utility differential:

$$\Delta V(\lambda) = -L\left[\frac{\theta}{\delta} - \frac{\Upsilon(\tau)}{\phi}\right]\left(\lambda - \frac{1}{2}\right), \tag{9}$$

where

$$\Upsilon(\tau) \equiv \frac{\tau[4\beta(3\beta+2) - (6\beta^2 + 6\beta + 1)\tau]}{2\beta(2\beta+1)^2},$$

which is positive because τ is smaller than τ_{trade}.

It follows immediately from (9) that $\lambda = 1/2$ is always a spatial equilibrium. This equilibrium is stable when θ exceeds $\delta\Upsilon(\tau)/\phi$. Otherwise, the manufacturing sector is concentrated into a single city. As a result, *when commuting costs steadily decrease, there is a transition from dispersion to agglomeration.* The intuition behind this result is straightforward. When θ is large, urban costs are sufficiently high to prevent the emergence of a big city. By contrast, there is agglomeration when θ is small because the gains from labor income and from more local varieties overcome the land market crowding effect. Note also that a high population density, a high labor productivity, or both makes agglomeration more likely because a larger city allows individuals to consume a wider range of varieties priced at a lower level.

Though very simple, the above model allows understanding the role played by commuting costs in shaping the space-economy. Consumers, who have a love for variety, are attracted by the city supplying the wider range of local varieties, which are cheaper than the imported varieties. By moving to this city, consumers increase the size of the local market, which makes local competition tougher. However, migration flows crowd out the land market and raise the urban costs borne by consumers residing in this city. Eventually, market clearing and labor mobility balance these various forces and select a spatial pattern involving either two small cities or one large city.

Note the difference with Krugman (1991) and the similarity with Helpman (1998): *Low transport costs are associated with the dispersion of activities.* Indeed, when τ is very small, we have $\Upsilon(\tau) \approx 0$, which implies $\delta\Upsilon(\tau) - \theta\phi < 0$. Consequently, firms and consumers are located in two small cities. This is because consumers have more or less the same access to the whole range of varieties, but obviate paying high urban costs through dispersion. This means that lowering transport costs induces the (partial) de-industrialization of large manufacturing cities and the relocation of manufactures in smaller cities or even in rural areas (Henderson 1997). On the contrary, when τ is large and smaller than τ_{trade}, $\Upsilon(\tau)$ takes on its largest value so that $\delta\Upsilon(\tau)/\phi$ is more likely to exceed θ. Indeed, when transport costs are high, the agglomeration of the manufacturing sector allows consumers to have direct access to all varieties at a low price while firms are able to better exploit scale economies. In other words, high transport costs are likely to be associated with the agglomeration of activities.

To sum up, a drop in the cost of shipping commodities fosters the spatial decentralization of jobs and production: *Krugman's prediction is thus reversed.* The intuition for this difference in results is easy to grasp. In the above model, urban costs rise when consumers join the larger city, which strengthens the dispersion force. Simultaneously, lowering transport costs facilitates intercity trade. Combining the two forces tells us why agglomeration or dispersion arises. By contrast, in the core-periphery model developed by Krugman (1991), the spatial concentration of workers does not generate any cost in the core. Furthermore, the dispersion force stems from the presence of immobile farmers who live in the two regions. This force gets weaker when farmers can be supplied at a lower cost. Consequently, manufacturing firms choose to locate in the same region to benefit from a larger market.

Finally, note that the catastrophic nature of the bifurcation obtained both here and in Krugman (1991) is an artifact due to the assumption of identical consumers. Once it is recognized that consumers are heterogeneous in their migration behavior, the agglomeration/dispersion process is gradual (Tabuchi and Thisse 2002). Therefore, the interest generated by the result of sudden urbanization is unwarranted.

2.3 Extensions

A. Agents' Heterogeneity. So far, agents have been assumed to be identical. However, it is well known that both firms and workers are heterogenous in productivity. Accounting for heterogeneous agents gives rise to a whole range of new issues (Behrens and Robert-Nicoud 2015). In what follows, we briefly discuss the well-documented existence of a wage premium: *Wages are higher in larger cities than in smaller cities* (Glaeser and Maré 2001; Combes et al. 2008; Combes and Gobillon 2015). We have seen that the agglomeration of workers leads firms to pay a higher wage to compensate workers for their higher urban costs. However, the wage premium may also occur because more talented and skilled workers sort into large cities, because large cities select the most productive firms, or both.

(i) We extend the framework developed in Section 2.1 by assuming that the economy involves skill-heterogeneous workers. A s-worker supplies s efficiency units of labor. Since the wage per efficiency unit of labor is given by Π_r/ϕ, the wage earned by a s-worker living in r is equal to $w_r^*(s) = s\Pi_r/\phi$. It is then readily verified that $\partial^2 V_r/\partial s \partial \lambda_r > 0$, which means that the single-crossing property holds. In this case, the incentive to live in the larger city is stronger for the high-skilled workers. Put differently, workers sort themselves across cities according to their efficiency.

When the space-economy involves a large and a small city, the high-skilled workers live in the large city because their earnings are sufficiently high to cover the higher urban costs. By contrast, the low-skilled workers will choose the small city where they earn lower wages, but bear lower urban costs. Under these circumstances, big cities or large urban regions are more productive than small cities because they house workers endowed with higher skills

(Behrens et al. 2014). In other words, the spatial concentration of skilled workers enhances the productivity of the economy at the expense of growing regional disparities. This result is in accordance with the empirical evidence provided by Moretti (2012) for the United States and by Combes et al. (2008) for France. In sum, *the unequal spatial distribution of human capital is one of the main causes of regional disparities.* Such a partition of the population may lead the those who are left behind to take a revenge via a wave of political populism that has strong territorial foundations (Rodríguez-Pose 2018).

(ii) How do firms spatially sort themselves according to their total factor productivity? The intuitive argument is easy to grasp. The less efficient firms seek protection against competition with the more efficient firms by locating in peripheral areas where competition is softer. By contrast, the more efficient firms choose to set up in the core because their higher productivity allows them to have a direct access to the larger market where they compete all together by setting lower prices but selling more (Baldwin and Okubo 2006; Okubo et al. 2010). In other words, *the selection of firms along the productivity line widens the gap between the core and the periphery.* However, when the spatial separation of markets ceases to be a sufficient protection against competition from the low-cost firms, high-cost firms also choose to set up in the larger market where they have a better access to a bigger pool of consumers.

B. The Impact of Local Factors. The spatial distribution of economic activities is also driven by regional advantages reflecting regional differences in technological capabilities or institutional quality.

(i) Commuting costs may be viewed as a proxy for the quality of *intra-regional* transport infrastructures. More generally, a lower of value of θ may be considered as an index whose value captures the quality of urban/regional governance through the various public services that enhance the quality of urban life. It is, therefore, worth looking at what the space-economy becomes when θ_A and θ_B are different. Assume that $\theta_A < \theta_B$. Everything else being equal, this implies that urban costs are lower in A than in B, thus allowing city A to host a larger population than B. Thus, when transport costs are low, the dispersed configuration now involves a larger city in A and a smaller one in B. However, this size difference does not imply spatial inequality because the bigger size of city A is associated with higher housing costs and longer commutes than in city B.

Market mechanisms can make the city with the less efficient infrastructures the core of economy. Assume, indeed, that $\theta_A = \theta$ and $\theta_B = \theta + \Theta$ with $\Theta > 0$. The utility differential becomes $\Delta V_\theta(\lambda) = \Delta V(\lambda) - \Theta(1-\lambda)/2$. In this case, city A attracts all activities for the range of transport cost values given in Section 2.1 since $\Delta V_\theta(1) = \Delta V(1)$. However, agglomeration may also occur in city B if $\Delta V_\theta(0) < 0$ or, equivalently, $\theta/\delta < \Upsilon(\tau)/\phi - \Theta$. Therefore, for a narrower range of transport cost values, the less efficient city may become the core of the global economy. Two comments are in order. First, *agglomeration can emerge everywhere, including areas with geographical disadvantage, through market mechanisms alone.* Second, when $\theta/\delta < \Upsilon(\tau)/\phi - \Theta$, there are two equilibria involving agglomeration, either in city A or in city B even though the former equilibrium Pareto-dominates the latter. In other words, the equilibrium spatial distribution need not coincide with the first best distribution.

(ii) It is well known that high housing costs limit the growth of large and efficient cities (Glaeser et al. 2006; Albouy and Ehrlich 2018). The "artificial scarcity" of land that stems from restrictive land use regulation, the provision of open spaces, or public policies that maintain the prices of agricultural products far above the international level, hurts new residents by reducing their welfare level or motivates a fraction of the city population to migrate away. This idea can be captured in a simple way by assuming that the opportunity cost of land $R_0 > 0$ is higher in, say, region A than in B. As a result, the land rents in the two cities are now given by the following expressions:

$$R_A^*(x) = \theta\left(\frac{\lambda_A L}{2} - \delta x\right) + R_0 \qquad R_B^*(x) = \theta\left(\frac{\lambda_B L}{2} - \delta x\right).$$

The indirect utility differential implies that city B is bigger than city A at the dispersed configuration. If, for whatever reason, the total factor productivity of city A is higher than city B's, making land more expensive in A prevents this city from exploiting its full potential through more concentration of activities. Indeed, some workers choose to stay put in the less efficient city where urban costs are lower. This illustrates how land use regulations, which often benefit the owners of existing plots and buildings, may restrict the growth of productive cities.

The above two examples illustrate the following key point: *Local policies may have perverse effects on the spatial organization of the economy.*

C. Technological Progress in Manufacturing. The bulk of NEG has focused on decreasing transport/trade costs. There is no doubt that the transport sector has benefited from huge productivity gains over the last two centuries. However, numerous other sectors have also experienced spectacular productivity gains. For this reason, Tabuchi et al. (2018) have revisited the core-periphery model by focusing on technological progress in the manufacturing sector. In addition, these authors assume that workers are imperfectly mobile. In other words, migration costs act here as the main dispersion force. In the core-periphery model, prices and nominal wages converge when transport costs decrease. Consequently, the real wage gap, hence the incentive to move, shrinks. When one region is slightly bigger than the other, technological progress in manufacturing reduces the labor marginal (respectively, fixed) requirement in the two regions, making the larger region more attractive by increasing wages and decreasing the prices of existing varieties (respectively, by increasing wages and the number of varieties) therein. Therefore, workers move to the larger region when the wider utility differential exceeds these workers' mobility costs. We may thus safely conclude that technological progress tends to exacerbate the difference between the two regions. As a result, *technological progress in manufacturing also favors agglomeration.* Like in the foregoing, urban costs slow down the spatial concentration of the manufacturing sector.

Since innovations often require skilled workers who are typically more mobile than unskilled workers, the core region is likely to host the most skilled workers. This sheds light on the following important question: Is interregional labor mobility the right solution to lessen regional disparities? Not necessarily. The standard pro-migration argument holds that interregional labor movement helps to limit the impacts of economic shocks and reduce regional disparities and structural unemployment in the long run (Blanchard and Katz 1992). There is no doubt that these forces are at work in the space-economy. However, the pro-migration argument disregards the fact that labor is probably the most differentiated production factor: Migrants are often the top-talent and most entrepreneurial individuals of their region. When the lagging region loses its best workers, those who are left behind are likely to be worse off. The redistributional consequences of mobility are, therefore, very different from what the standard model predicts.

D. The Bell-Shaped Curve of Spatial Development. Historically, it is well known that both transport and commuting costs have fallen at an unprecedented pace. Therefore, what matters is the relative evolution of these two types of spatial costs. For a long time, high transport costs have been the main impediment to trade. Nowadays, within developed countries, the cost of shipping commodities has reached a level that is probably lower than commuting costs, which remain relatively high. As a consequence, the main dispersion force no longer lies in the cost of supplying distant markets, but in the level of urban costs. Under these circumstances, one may speculate that, though economic integration would have initially fostered a more intensive agglomeration of economic activities, its continuation is liable to generate a redeployment of activities that could lead to a kind of geographical evening-out. Medium-sized and small 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. When they combine cheaper labor and land with a business-friendly local culture, lagging regions have a cost advantage that offsets their lack of agglomeration effects. Hence one may expect the process of spatial development to unfold according to a *bell-shaped curve*. To be precise, the relationship between economic integration and spatial inequality is not monotone: While the first stages of economic integration exacerbate regional disparities, additional integration starts undoing them (for a more detailed discussion of the bell-shaped curve, see Fujita and Thisse 2013).

2.4 The Decentralization of Jobs Within Cities

As seen above, globalization could well challenge the supremacy of large cities, the reason being that the escalation of urban costs would shift employment from large monocentric cities to small cities where these costs are lower. However, this argument relies on the assumption that cities have a monocentric morphology. The main point we wish to stress here is that decentralizing the production of goods in *secondary business districts* (SBD) may allow large cities to retain a high share of firms and jobs. Under these circumstances, firms are able to pay lower wages while retaining most of the benefits generated by large urban agglomerations. For example, Timothy and Wheaton (2001) report substantial variations in wages according to intra-urban location (15% higher in central

Boston than in outlying work zones, 18% between central Minneapolis and the fringe counties). As they enjoy living on larger plots and/or move along with firms, consumers may also want to live in suburbia. Consequently, the creation of subcenters within a city, that is, the formation of a *polycentric city*, appears to be a natural way to alleviate the burden of urban costs.

For the redeployment of activities in a polycentric pattern to happen, firms set up in SBDs must be able to maintain a good access to the main urban center, which requires low communication costs. For example, a large share of the business services consumed by US firms located in suburbia are supplied in city centers. By focusing on urban and communication costs, we recognize that both agglomeration and dispersion may take two quite different forms because they are now compounded by the *centralization* or *decentralization* of activities within the same city. Such a distinction is crucial for understanding the interactions between cities and trade.

A. Polycentric Cities. We extend the above model by allowing manufacturing firms to locate in the CBD or to form a SBD on each side of the CBD (Cavailhès et al. 2007). Both the CBD and the SBDs are surrounded by residential areas occupied by consumers. Because the higher-order services are still provided in the CBD, firms established in a SBD must incur a communication cost $K > 0$ so that the profit of a firm located in a SBD is given by $\pi_r - K$ whereas π_r is the profit of a firm established in CBD. In what follows, the superscript C is used to describe variables related to the CBD, whereas S describes the variables associated with a SBD.

Denote by y_r the right endpoint of the area formed by residents working in the CBD and by z_r the right endpoint of the residential area on the right-hand side of the SBD, which is also the outer limit of city r. Let x_r^S be the center of the SBD in city r. It is easy to show that these points are given by

$$y_r = \frac{\sigma_r \lambda_r L}{2\delta} \qquad x_r^S = \frac{(1+\sigma_r)\lambda_r L}{4\delta}, \tag{10}$$

where $\sigma_r < 1$ is the share of jobs located in the city r-CBD.

Individuals choose their workplace (CBD or SBD) and their residential location, given the land rents and wages in the CBD (w_r) and the SBD (w_r^S). The wage wedge between the CBD and a SBD is given by

$$w_r^C - w_r^S = \theta(2y_r - x_r^S) = \frac{\theta}{\delta}\frac{3\sigma_r - 1}{4}\lambda_r L, \tag{11}$$

where we have used the expressions for y_r and x_r^S given in (10). In other words, the difference in the wages paid in the CBD and in the SBD compensates exactly the worker for the difference in the corresponding commuting costs. Moreover, the wage wedge is positive as long as $\sigma_r > 1/3$, that is, the size of the CBD exceeds the size of each SBD. Note also that a larger population in city r raises the wage wedge. Indeed, as the average commuting cost rises, firms located in the CBD must pay a higher wage to their workers.

At each workplace (CBD or SBD), the equilibrium wages are determined by a bidding process in which firms compete for workers by offering them higher wages until no firm earns positive profits. Hence, the equilibrium wages are related through the following expressions: $w_r^{C*} = w_r^*$ and $w_r^{S*} = w_r^* - K$. Substituting w_r^{C*} and w_r^{S*} into (11) and solving with respect to σ_r yields the equilibrium relative size of the CBD in city r:

$$\sigma_r^* = \min\left\{\frac{1}{3} + \frac{4\delta K}{3\theta\phi\lambda_r L}, 1\right\},$$

which always exceeds 1/3. Clearly, the city is polycentric ($\sigma_r^* < 1$) if and only if

$$K < \frac{\theta\phi\lambda_r L}{2\delta}.$$

The higher the communication costs, the lower the commuting cost, or both, the larger the CBD. At the limit, when communication keep rising, both SBDs shrink smoothly and the city becomes monocentric. Note also that a bigger city is more likely to be polycentric.

B. The Emergence of Polycentric Cities. The utility differential between cities now depends on the degree of job decentralization within each city. The indirect utility of an individual working in the CBD is still given by (8) but the urban costs (5) are replaced by the following expression:

$$UC_r \equiv \frac{\theta\lambda_r L}{2\delta}\sigma_r^*.$$

Everything else equal, urban costs take on lower values when jobs are decentralized into the SBDs. It thus follows that the existence of SBDs allows the large cities to maintain their primacy.

The utility differential (9) becomes

$$\Delta V(\lambda) = -L\left[\frac{\theta}{3\delta} - \frac{\Upsilon(\tau)}{\phi}\right]\left(\lambda - \frac{1}{2}\right)$$

when both cities are polycentric and

$$\Delta V(\lambda) \equiv -2L\left[\frac{2\theta}{3\delta} - \frac{\Upsilon(\tau)}{\phi}\right]\lambda + L\left[\frac{\theta}{\delta} - \frac{\Upsilon(\tau)}{\phi} - \frac{4K}{3}\right]$$

when only one city is polycentric ($\sigma_1^* < \sigma_2^* = 1$).

The spatial economy displays a richer set of stable equilibrium configurations, which now also includes partial agglomeration with one large polycentric city and a small monocentric city, as well as agglomeration within a single polycentric (monocentric) city and dispersion with two identical polycentric (monocentric) cities. Once communication costs are low enough, the economy traces out the following path when the ration θ/δ steadily decreases. By inducing high urban costs, a high θ/δ-ratio leads to the dispersion and decentralization of jobs, that is, the economy involves two polycentric cities. When δ gets higher or θ lower, urban costs decrease sufficiently for the centralization of jobs within one city to emerge at the market outcome. However, urban costs remain high enough for the equilibrium to involve two cities having different sizes and structures. Last, when the θ/δ-ratio takes on very low values, urban costs become small enough for saving the cost of shipping the manufactured good through the existence of a single large city.

The multiplicity of stable equilibria also has an important implication that has been overlooked in the literature: *Different types of spatial patterns may coexist under identical technological and economic conditions.* It should be no surprise, therefore, to observe different types of urban systems in the real world.

3 Cities and Services

In the models presented in Section 2, consumers have access to the entire range of produced varieties. However, cities supply a rising share of local consumption services. What distinguishes service cities from manufacturing cities is that the cost of shipping local services are prohibitive. Consequently, consumers have access only to the varieties produced in the city in which they live.

3.1 The Formation of Consumption Cities

Consumer preferences are still given by (2), except that the set of available varieties in city r is now n_r instead of n. The profits earned by a city r-service firm are given by

$$\pi_r = p_r q_r(p_r)\lambda_r L - w_r\phi.$$

Since service firms compete only on their local market, the equilibrium price of a city r-variety is obtained by setting $\lambda_s = 0$ in (7):

$$p_r^* = \frac{\beta}{2\beta + 1}, \tag{12}$$

which is independent of the city size and the same in the two cities. Observe that a stronger love for variety yields a higher market price because service firms have more market power. Therefore, the consumer surplus generated by any single variety does not vary between cities ($S_A = S_B = S$). Since the value of S does not play any role in the analysis undertaken below, we may set $S = 1$. Thus, the total surplus is equal to the number $n_r = \lambda_r L/\phi$ of locally produced varieties, which increases with the size of the city. In other words, *consumers living in larger cities have access to a wider array of untradable services.*

The urban labor markets being local, the equilibrium wage paid by firms established in city r is equal to

$$w_r^* = \frac{\mathbf{p}}{\phi}\lambda_r L \qquad \text{with } \mathbf{p} \equiv \frac{\beta}{(2\beta+1)^2}. \tag{13}$$

Consequently, *wages are higher in larger cities.*

Replacing each term of V_r by its new expression leads to the following utility differential:

$$\Delta V(\lambda) = -L\left[\frac{\theta}{\delta} - \frac{2(1+\mathbf{p})}{\phi}\right]\left(\lambda - \frac{1}{2}\right). \tag{14}$$

As in Section 2, the symmetric pattern ($\lambda^* = 1/2$) is always a spatial equilibrium. However, when $\theta/\delta < 2(1+\mathbf{p})/\phi$, this equilibrium becomes unstable because the utility differential is positive for all values of λ. The market outcome therefore involves a single large city accommodating all consumers ($\lambda^* = 1$). By being agglomerated in a single city, consumers have access to all varieties. Thus, *even in the absence of trade, consumers and firms may choose to be agglomerated within a single large city.* This is so when (i) commuting costs are low, (ii) the population density ($1/\delta$) is high, and (iii) the range of local services is wide. Furthermore, low fixed costs favor the entry of additional firms, which widens the range of varieties and increases consumers' utility. As a consequence, labor-saving innovations push toward the concentration of services in large cities.

By contrast, when $\theta/\delta > 2(1+\mathbf{p})/\phi$, the symmetric equilibrium is stable. This is because the gains from variety do not compensate consumers for the higher urban costs they would bear in the large city. In this case, instead of seeking variety, consumers aim to reduce urban costs. The population is thus equally dispersed.

To sum up, when commuting costs steadily decrease a service economy shifts from dispersion to agglomeration because the latter allows individuals to consume all services and to earn higher wages. In other words, *urbanization may arise in the absence of industrialization.* The existence of large consumption cities in several developing countries is evidence that urbanization may arise in the absence of manufacturing sectors. For example, Gollin et al. (2016) find that countries that heavily depend on natural resources are associated with the emergence of "consumption cities," which produce mainly services.

3.2 Extensions

A. Spatial Externalities. Cities generate costs and benefits that arise from non-markets interactions (Duranton and Puga 2004). We focus here on *agglomeration economies* that affect positively the productivity of firms located in the same city because spillovers are typically subject to strong distance-decay effects (Carlino and Kerr 2015). More precisely, we assume that ϕ_r is such that $1/\phi_r = L_r^{\varepsilon}/\phi$ where $\varepsilon > 0$ is the elasticity of agglomeration economies with respect to population size. Likewise, we consider the negative externality associated with *traffic congestion* by assuming that the unit commuting cost is given by $\theta_r = \theta L_r^{\kappa}$ with $\kappa > 0$. The objective is to study the implications of these externalities for the inter-city distribution of activities.

The labor productivity of city r is such that $n_r q_r / L_r = L_r^{\varepsilon}\mathbf{p}/\phi$, the equilibrium wage is $w_r^* = \mathbf{p}L_r^{1+\varepsilon}/\phi$, while urban costs are equal to $\theta L_r^{1+\kappa}/2$. Plugging these new terms in V_r leads to the following utility differential:

$$\Delta V(\lambda) = -\frac{L}{2}\left(\frac{\theta}{\delta}\left\{(\lambda L)^{\kappa}\lambda - [(1-\lambda)L]^{\kappa}(1-\lambda)\right\} - \frac{2(1+\mathbf{p})}{\phi}\left\{(\lambda L)^{\varepsilon}\lambda - [(1-\lambda)L]^{\varepsilon}(1-\lambda)\right\}\right).$$

Agglomeration is a stable equilibrium if and only if $\Delta V(1) > 0$. This condition amounts to

$$\theta < 2\delta\frac{1+\mathbf{p}}{\phi}L^{\varepsilon-\kappa} \equiv \theta_A.$$

On the other hand, when

$$\theta > \theta_D \equiv 2\delta\frac{\varepsilon+1}{\kappa+1}\frac{1+\mathbf{p}}{\phi}\left(\frac{L}{2}\right)^{\varepsilon-\kappa}$$

holds, the symmetric equilibrium is stable because the slope of $\Delta V(\lambda)$ is negative at $\lambda = 1/2$.

Note that $\theta_A < \theta_D$ when agglomeration economies are stronger than the congestion effect ($\varepsilon > \kappa$). In this case, a larger city and a smaller city emerge in equilibrium when $\theta_A < \theta < \theta_D$. In other words, spatial externalities foster the emergence of an urban hierarchy. By contrast, when the congestion externalities are strong enough ($\kappa > \varepsilon$), we have $\theta_D < \theta < \theta_A$ so that both agglomeration ($\lambda = 1$) and dispersion ($\lambda = 1/2$) are stable equilibria. In sum, *spatial externalities that percolate within cities affect the distribution of activities between cities* (Henderson 1988).

B. Exogenous Amenities. Natural and historical amenities are typical examples of untradable goods that consumers enjoy (Koster and Rouwendal 2017). The literature on migration shows that amenities play a significant role in the location of consumers (Kahn and Walsh 2015). Since consumers prefer more amenities than less, it is reasonable to consider a preference structure similar to the one used in models of vertical product differentiation, that is, $U_r^a = a_r U(q_0, q_i)$ where a_r is the amenity level in city r. In this case, the indirect utility function is given by $V_r^a = a_r V_r(\lambda_r)$. In what follows, we consider two opposite, but relevant, cases. In the former, consumers have different tastes while the amenity level is the same in the two cities. In the latter, consumers have different incomes while the amenity level varies across city locations.

(i) Assume that the amenity level a is the same in both cities. Since consumers are heterogeneous in their perception of cities, the matching value between a consumer and city r is described as the realization of a random variable ε_r, where the variables ε_r are i.i.d. across consumers according to the Gumbel distribution with zero mean and a variance equal to $\pi^2 v^2/6$. A larger value of ν means that consumers are more heterogeneous in taste.

The probability that a consumer chooses city r is given by the binary logit:

$$\mathbb{P}(\lambda) = \frac{\exp[aV_r(\lambda)/v]}{\exp[aV_A(\lambda)/v] + \exp[aV_B(\lambda)/v]}. \tag{15}$$

A spatial equilibrium is now such that the flow of in-migrants is equal to the flow of out-migrants: $(1-\lambda)\mathbb{P}(\lambda) = \lambda\mathbb{P}(1-\lambda)$. Taking the logarithm of this expression and plugging (15) yields the equilibrium condition:

$$\Delta V(\lambda) - \frac{v}{a} \log \frac{\lambda}{1-\lambda} = 0, \tag{16}$$

where $\Delta V(\lambda)$ is given by (14). The symmetric equilibrium is stable if and only if $\theta/\delta > 2(1+\mathbf{p})/\phi - 4v/a$ holds because the slope of the left-hand side of (16) is negative at $\lambda = 1/2$. Hence, the more heterogeneous the population of consumers ($\nu \upharpoonleft$), the more likely the dispersed configuration. In other words, *taste heterogeneity acts as a dispersion force.*

When commuting costs are such that $\theta/\delta < 2(1+\mathbf{p})/\phi - 4v/a$, the spatial equilibrium involves a larger city and a smaller city. In other words, there is *partial agglomeration*, namely $1/2 < \lambda^* < 1$. Furthermore, since $\mathrm{d}\lambda^*/\mathrm{d}\theta < 0$ and $\mathrm{d}\lambda^*/\mathrm{da} > 0$, *lower commuting costs or a higher amenity level leads to a wider range of services in the bigger* *city*. Indeed, the gains associated with agglomeration are magnified with the amenity level because a acts as a multiplication constant of $\Delta V(\lambda)$.

(ii) We now consider the case of consumers heterogeneous in incomes w. In a "featureless" city, there is spatial segregation, namely the distance between consumers' residential locations is perfectly correlated to their income gap (Fujita 1989). However, in a "featureful" city the relative location of different income groups is also affected by the spatial pattern of urban amenities (Brueckner et al. 1999). For example, Cuberes et al. (2019) find no evidence of a strong relationship between income and household distance to the center of English cities.

We assume that the amenity level $a_r(x)$ varies across locations within city r. Furthermore, the unit commuting cost increases with income because the opportunity cost of time rises with income. This cost is now given by θw rather than θ. Consequently, the indirect utility of a w-consumer located at x in city r is such that

$$V_r^a(x, w) = a_r(x)[(1 - \theta x)w - R_r(x)/\delta + n_r + q_0^*].$$

By definition, the bid rent $\Psi_r(x, w)$ is such that the w-consumers are indifferent across locations in city r. Therefore, $\Psi_r(x, w)$ must solve the equilibrium condition $\mathrm{d}V_r^a/\mathrm{d}x = 0$ (Fujita 1989). This solution can be shown to be given by the following expression:

$$\Psi_r(x, w) = \delta \left[(1 - \theta x)w + n_r + q_0^* + \frac{k_r(w)}{a_r(x)}\right],$$

where $k_r(w)$ is a constant that depends on w.

The problem consists in assigning consumers having particular incomes to specific locations within city r under the form of a mapping $w(x)$ that specifies the income of consumers assigned to x. Since land at x is allocated to the highest bidder, the equilibrium income mapping $w^*(x)$ must solve the utility-maximizing condition $\partial\Psi_r/\partial w = 0$, while the second-order condition implies $\partial^2\Psi_r/\partial w^2 < 0$. Totally differentiating $\partial\Psi_r/\partial w = 0$ with respect to x yields the equilibrium income mapping as a solution to the following differential equation:

$$\frac{\mathrm{d}w_r^*}{\mathrm{d}x} = -\left[\frac{\partial^2\Psi_r(x, w)}{\partial w^2}\right]_{w^*(x)}^{-1} \cdot \left.\frac{\partial^2\Psi_r(x, w)}{\partial w \partial x}\right|_{w^*(x)}.$$

Since $\partial^2\Psi_r/\partial w^2 < 0$, $\partial\Psi_r/\partial w\partial x$ and $\mathrm{d}w^*(x)/\mathrm{d}x$ have the same sign. Hence the interaction between the amenity and commuting cost functions determines the social stratification of the city through the behavior of the function $\partial^2\Psi_r/\partial w\partial x$.

It also follows from $\partial\Psi_r/\partial w = 0$ that $\partial k_r(w)/\partial w = -a_r(x)(1 - \theta x)$. This equality implies

$$\frac{\partial^2\Psi_r}{\partial w\partial x} = (1 - \theta x)\left[\frac{1}{a_r(x)}\frac{\partial a_r(x)}{\partial x} + \frac{\theta}{1 - \theta x}\right].$$

Set $\Lambda_r(x) \equiv a_r(x)(1 - \theta x)$, a magnitude that has the nature of an hedonic locational index. Differentiating $\Lambda_r(x)$ with respect to x shows that $\mathrm{d}\Lambda_r(x)/\mathrm{d}x$ and $\partial^2\Psi_r/\partial w\partial x$ have the same sign. The income sorting of consumers within city r is thus driven by the behavior of $\Lambda_r(x)$ with respect to x. When the equation $\partial^2\Psi_r/\partial w\partial x = 0$

have several solutions, there is *imperfect sorting* in that greater income differences are no longer mapped into more spatial separation. The affluent locate where the hedonic locational index is maximized, whereas the poor reside in locations with the lowest hedonic locational index. Inside the remaining areas, various residential patterns may emerge. For example, the middle class may be split into two spatially separated neighborhoods with the poor in between, while the rich live near the city center.

How consumers choose a city is determined by combining the two urban spaces within a location set in which the two CBDs are separated by a given distance. Repeating the above argument shows that consumers select locations, which are both a city location and a residential location in this city. We may conclude that *amenities affect agglomeration and dispersion forces.*

4 The Industrial Structure of Cities

We now take a broader perspective by considering an economy with a manufacturing sector supplying a tradable good and a sector producing a untradable service for local consumption. In this economy, labor is perfectly mobile between cities *and* sectors. Such a two-sector setting generates a richer set of results than models with a single industry as the objective is to determine the sectorial and spatial structure of the economy.

4.1 Manufactured Goods and Services

The manufactured good is denoted by 1 and the service by 2. The utility derived from consuming q_i units of a variety i of good $j = 1, 2$ is given by (1). The parameters associated with the utility arising from consuming one variety of the manufactured product or of the consumption service are identical. This assumption, which is made for analytical simplicity, does not affect qualitatively the properties of the spatial equilibria. Indeed, since good 1-varieties are available everywhere at the same price, the consumer surplus generated by the consumption of the manufactured good is the same regardless of the city where consumers live. Furthermore, the profits earned by the manufacturing firms are also the same regardless of the city where firms are located. Consequently, the equilibrium values of the consumer surplus and wage associated with good 1 do not play any role in workers' migration decision. Hence, assuming that the parameters of (1) are the same for goods 1 and 2 entails no significant loss of generality.

Preferences involve two goods and are given by

$$U(q_0; q_{ij}) = \sum_{j=1,2} \int_0^{n_j} q_{ij} \mathrm{d}i - \frac{\beta n_j}{2(n_1 + n_2)} \int_0^{n_j} q_{ij}^2 \mathrm{d}i - \frac{1}{2(n_1 + n_2)} \int_0^{n_j} q_{ij} \left(\int_0^{n_j} q_{kj} \mathrm{d}k \right) \mathrm{d}i + q_0, \tag{17}$$

where q_{ij} is the quantity of variety $i \in [0, n_j]$ of good $j = 1, 2$. Since $\beta > 0$, (17) encapsulates both a preference for diversity between goods 1 and 2, as well as a preference for variety across varieties of the same good. Since the second term of (17) is weighted by the ratio $n_j/(n_1 + n_2)$, the intensity of the love for variety varies between services and the manufactured good. In particular, a good or service supplied as a small range of varieties has more impact on the consumer well-being than a good or service made available through a large array of varieties.

Good 1 is assumed to be freely tradable, so that the total number n_1 of good 1-varieties is available in both cities. By contrast, a consumer living in city $r = A, B$ has access to the number n_{2r} of good 2-varieties supplied in city r. In a one-sector economy, the distribution of firms and workers is a priori undetermined when transport costs are zero. As shown below, the presence of the service sector is sufficient for a specific distribution of the manufactured sector to emerge.

Setting $\tau = 0$ in (7) yields the equilibrium price (12) of good 1, which is now constant and the same in both cities. This implies firms pay the same wage w_1 to all their workers because they earn the same operating profits regardless of their locations.

Let λ_{ir} be the endogenous number of consumers working in sector $i = 1, 2$ and living city $r = A, B$. Labor market-clearing implies

$$n_1 = \frac{(\lambda_{1A} + \lambda_{1B})L}{\phi_1} \qquad n_{2r} = \frac{\lambda_{2r}L}{\phi_2}, \quad \text{with } r = A, B.$$

Since labor is mobile between sectors, in equilibrium it must be that $w_{1r} = w_{2r} \equiv w_r$ for $r = A, B$. Setting $\lambda_r \equiv \lambda_{1r} + \lambda_{2r}$ for the population residing in city r, the budget constraint of a consumer residing in this city is as follows:

$$n_1 p_{1r} q_{1r} + n_{2r} p_{2r} q_{2r} + \frac{\theta}{\delta} \frac{\lambda_r L}{2} + q_0 = \bar{q}_0 + w_r.$$

It can readily be shown that the individual demand for a good i-variety in city r is given by

$$q_{1r} = \left(\frac{1}{1+\beta} - \frac{p_1^*}{\beta} + \frac{\bar{p}_1}{\beta(\beta+1)} \right) \left(1 + \frac{n_{2r}}{n_1} \right), \tag{18}$$

$$q_{2r} = \left(\frac{1}{1+\beta} - \frac{p_{2r}}{\beta} + \frac{\bar{p}_{2r}}{\beta(\beta+1)} \right) \left(1 + \frac{n_1}{n_{2r}} \right). \tag{19}$$

Although this demand system involves no income effect, it displays a rich pattern of substitution via the relative number of varieties. More specifically, when the number of good i-varieties available in city r increases, the individual demands for good j-varieties are shifted upward because good j becomes relatively more attractive. Furthermore, the size and distribution of the service sector (n_{2r}) affects individual demands for the manufactured good in each city (see (19)). In particular, a growing service sector has a positive impact on the demand for the tradable good.

In addition, the demand system (18)-(19) highlights one of the main differences with the one-sector model, that is, the number of varieties n_i produced in industry i is endogenous because it depends on the mobility of workers between sectors. Finally, note that the size of the manufacturing sector (n_1) affects the individual demand for services in each city and, therefore, the spatial distribution of this sector. In contrast, the distribution of manufactures has no direct impact on individual demands for good 1 because trading this good is costless. This suggests that these firms are indifferent between locations. Nevertheless, they are not because their workers are attracted by cities supplying a wider range of services.

Let

$$\pi_{1r} \equiv p_1[q_{1A}(p_1)\lambda_A + q_{1B}(p_1)\lambda_B]L - \phi_1 w_r$$

be the profits earned by a manufacturing firm established in city r. As in Section 2, when choosing its own price, each firm treats parametrically the wage w_r, as well as the average prices $\bar{p}_{1A}$ and $\bar{p}_{1B}$.

Profits being zero in equilibrium, it follows from (18) that the wage paid by a manufacturing firm is equal to

$$w_1 = \mathbf{p} \sum_{r=A,B} \frac{n_r \lambda_r L}{n_1 \phi_1}, \tag{20}$$

where $n_r \equiv n_1 + n_{2r}$ and $\mathbf{p}$ is defined in (13).

The profits made by a service firm set up in city r are given by

$$\pi_{2r} = p_{2r} q_{2r} \lambda_r L - \phi_2 w_r,$$

where p_{2r} is the price quoted by such a firm. Since the equilibrium price of a good 2-variety is given by (12), the equilibrium wage paid by service firms located in city r is given by

$$w_{2r} = \mathbf{p} \frac{n_r \lambda_r L}{n_{2r} \phi_2}, \tag{21}$$

which varies with the size ($\lambda_r L$) and the sectorial mix (n_r/n_{2r}) of city r. Note that *the service sector is never agglomerated*, for otherwise w^* would becomes arbitrarily large (set $n_{2r} = 0$ in (21)).

Since the individual consumer surplus $S = 1$, the welfare of a consumer working in sector i and living in city r is given by the following expression:

$$V_{ir} = n_1 + n_{2r} + w - \frac{\theta}{\delta} \frac{\lambda_r L}{2},$$

which shows how consumers' well-being depends on the spatial and sectorial distribution of jobs.

It follows from this expression that, in equilibrium, the urban cost differential is exactly compensated by the difference in the number of untradable services supplied in each city. In other words, *consumers choose to live in a larger city where they bear higher urban costs because they have access to a wider array of local services.*

A *spatial-sectorial equilibrium* arises when no worker has an incentive to change place and/or to switch job. How to describe the mobility of workers between regions and sectors? Tabuchi and Thisse (2006) treat the choices of jobs and locations in a symmetric way and borrow the myopic evolutionary dynamics used in NEG to study the mobility of workers across several regions (Fujita et al. 1999a):

$$\dot{\lambda}_{ir} = \lambda_{ir} \left(V_{ir} - \bar{V} \right),$$

where $\bar{V} = \Sigma_i \Sigma_r \lambda_{ir} V_{ir}$ is the average utility in the global economy.

The perfect mobility of labor between sectors implies factor price equalization between cities and sectors, that is, $w_1 = w_{2A} = w_{2B}$. It then follows from (20) and (21) that the labor force is equally split between the two sectors

($\lambda_1^* = \lambda_2^* = 1/2$). Note that this equality does not mean that workers and sectors are equally distributed between the two cities.

The utility differential can be shown to be given by

$$V_{ir} - \bar{V} = \lambda_s L \left[\frac{2\delta - \theta\phi_2}{\delta\phi_2} \left(\lambda_{2r} - \frac{1}{4} \right) - \frac{\theta}{\delta} \left(\lambda_{1r} - \frac{1}{4} \right) \right],$$

where λ_s is the city s-population with $s \neq r$. Solving this system of equations shows that there are two candidate equilibria (up to a permutation between A and B):

$$\lambda_{1A}^* = \lambda_{2A}^* = 1/4 \tag{22}$$

and

$$\lambda_{1A}^* = \frac{1}{4} + \frac{(2\delta - \theta\phi_2)\sqrt{\Delta}}{4\theta\phi_1\phi_2} \qquad \lambda_{2A}^* = \frac{1}{4} + \frac{\sqrt{\Delta}}{4\phi_1}, \tag{23}$$

where $\Delta \equiv \phi_1^2 + 2\phi_1\phi_2 - 2\phi_1\phi_2^2\theta/\delta$. As in Section 2, the symmetric pattern (22), which involves two cities having the same size and the same industrial mix, is always a spatial-sectorial equilibrium. On the other hand, the asymmetric configuration (23) is a stable equilibrium if and only if $\Delta > 0$, that is,

$$\theta < \delta \left(\frac{\phi_1}{2\phi_2^2} + \frac{1}{\phi_2} \right).$$

Therefore, commuting costs must be sufficiently low, or the productivity of the service sector must be sufficiently high, for a large city (A) and a small city (B) to coexist. Moreover,

$$\lambda_A^* - \lambda_B^* = \frac{\delta\sqrt{\Delta}}{\phi_1\phi_2\theta} \geq 0$$

implies that *lowering commuting costs gradually enlarges the population gap between the two cities.* Though workers are identical, there is no catastrophic bifurcation: Small changes in commuting costs generate small changes in the location *and* the composition of economic activities. In other words, accounting for the possibility of changing jobs is sufficient to smooth out the process of migration.

In the symmetric configuration (22), the total number of good i-varieties is given by $n_i^* = 1/(2\phi_i)$ for $i = 1, 2$. Hence, the industrial mix $n_1^*/n_2^* = \phi_2/\phi_1$ of the economy depends on the relative productivity of labor in the two sectors. By contrast, when the asymmetric configuration (23) prevails, *cities differ not only in size but also in their industrial mix* through their respective numbers of service varieties.

Last, as said above, the existence of untradable services selects a well-defined distribution for the footloose industry 1. More precisely, except for sufficiently high commuting costs, *the untradable sector acts as a centripetal force* that leads to the partial agglomeration of the manufacturing sector.

A. The Export Base Theory Revisited. Focusing on an industrial mix that combines a tradable good and an untradable service allow us to revisit the export base theory grounded in the assumption that the urban economy can be divided into two very broad sectors, that is, a basic sector whose fortunes depends largely in external factors and a nonbasic sector that depends on local factors. The tenet of this theory holds that the basic sector is the prime cause for local economic growth (Tiebout 1956).

Using the equilibrium condition $V_{iA}^* = V_{iB}^*$, it can readily be shown that

$$\lambda_A^* = \frac{1}{2}\frac{\delta - \phi_2\theta}{2\delta - \phi_2\theta} + \frac{2\delta}{2\delta - \phi_2\theta}\lambda_{1A}^*,$$

where $2\delta/(2\delta - \theta\phi_2) > 1$ is the "regional multiplier". Therefore, a shock that makes the basic sector larger (λ_{1A}^*) boosts a more than proportionate growth of the city size (λ_A^*) by attracting a larger number of service firms. However, a larger nonbasic/service sector also leads to the expansion of the manufacturing sector, and thus the nonbasic sector can also serve as an engine of urban growth.

Observe also that, the impact of the service sector on total employment is higher in the larger city because this sector is relatively more concentrated in city A than in city B. Indeed, the equilibrium condition $V_{iA} = V_{iB}$ implies

$$\lambda_A^* = \frac{\phi_2\theta - \delta}{2\phi_2\theta} + \frac{2\delta}{\phi_2\theta}\lambda_{2A}^*,$$

so that the regional multiplier of the nonbasic sector exceeds of the regional multiplier of the basic sector when $\theta > \delta/\phi_2$, an inequality that is more likely to hold when the productivity in the service sector is high.

To sum up, *manufacturing sectors are not necessarily a more powerful engine of urban growth than service sectors.*

B. Urban Hierarchy. As long as $\theta < 2\delta/\phi_2$ holds, the larger city supplies a larger array of varieties of each good than the smaller city ($\lambda_{1A}^* > 1/4 > \lambda_{1B}^*$ and $\lambda_{2A}^* > 1/4 > \lambda_{2B}^*$). In this case, the urban system displays a *Christaller-like hierarchy*: By supplying a wider range of services, city A attracts more consumers than city B. Though the demand for the manufactured good is higher in A (see (18)), this city does not attract all manufactured workers because this good is shipped at zero cost. Thus, the process of circular causation comes to an end before reaching the full agglomeration of manufactures. Note, however, that the population gap between the two cities grows when the service sector becomes more productive.

C. Relative Comparative Advantage. When $\theta > 2\delta/\phi_2$ holds, the larger city has a larger labor share in the service sector, whereas the smaller city has a larger labor share in the manufacturing sector:

$$\frac{\lambda_{1B}^*}{\lambda_B^*} > \frac{1}{2} > \frac{\lambda_{1A}^*}{\lambda_A^*} \qquad \frac{\lambda_{2A}^*}{\lambda_A^*} > \frac{1}{2} > \frac{\lambda_{2B}^*}{\lambda_B^*}.$$

In other words, the urban system involves relatively specialized cities in that *the manufacturing sector is relatively less concentrated than services.* This is because the size advantage associated with the larger city no longer compensates the remaining manufacturing workers for the higher urban costs they would bear therein. This result is consistent with the theory of relative comparative advantage: The larger city has a comparative advantage in untradables because it has a larger local market, while the smaller city' comparative advantage lies in its lower urban costs. Note that cities' comparative advantage are not exogenous. Rather, they emerge as the outcome of market interaction and labor mobility. This relative specialization also agrees with the empirical evidence: Services have a taste for large cities that manufactures no longer have.

4.2 The Impact of Trade on the Structure of Cities

The setting developed above may be used to shed light on the impact of international trade on the distribution of activities between cities and sectors in the case of a small open economy which imports additional varieties of the manufactured good from the rest of the world. Denote by n_1^d (n_1^ω) the endogenous (exogenous) number of domestic (foreign) varieties of this good. Therefore, the total number of good 1-varieties available in the small open economy is $n_1 = n_1^d + n_1^\omega$ where $n_1^d = \lambda_1 L/\phi_1$.

The profit earned by a foreign firm exporting to the small economy is given by

$$(p_1^\omega - T)\left(q_{1A}^\omega \lambda_A L + q_{12}^\omega \lambda_B L\right),$$

where p_1^ω is the common price of a foreign variety in each city and $T > 0$ the level of international trade costs, while q_{1r}^f is the individual demand for this variety in city $r = 1, 2$:

$$q_{1r}^\omega = \left(\frac{1}{1+\beta} - \frac{p_1^\omega}{\beta} + \frac{\bar{p}_1}{\beta(\beta+1)}\right)\left(1 + \frac{n_{2r}}{n_1}\right) \qquad \text{with} \quad \bar{p}_1 = \frac{n_1 p_1 + n_1^\omega p_1^\omega}{n_1}.$$

The difference with (18) and (19) is that the average price $\bar{p}_1$ depends also on the prices of the foreign varieties.

The profit-maximizing conditions show that the consumer prices of the manufactured good charged by the foreign and domestic firms are, respectively, given by

$$p_1^{\omega *} = p_1^{d*} + \frac{T}{2} \qquad p_1^{d*} = \frac{\beta}{2\beta+1}\Phi(\lambda_1),$$

where

$$\Phi(\lambda_1) \equiv 1 + \frac{T}{2\beta}\frac{n_1^\omega}{n_1^d + n_1^\omega}. \tag{24}$$

Observe that the equilibrium prices $p_1^{\omega *}$ and p_1^{d*} of the manufactured good capture the pro-competitive effects associated with a larger number of imported varieties ($n_1^\omega \uparrow$), lower trade barriers ($T \downarrow$), and higher labor productivity in the manufacturing sector ($\phi_1 \downarrow$). When $n_1^\omega = 0$, we fall back on the expression $p_1^{d*} = p_1^*$ obtained above.

We have seen that $w_1^* = w_{2A}^* = w_{2B}^*$ holds in equilibrium. The zero-profit condition then implies that the equilibrium allocation of labor between the two sectors is given by the solution to the equation:

$$\lambda_1(\Phi) = \frac{1 - n_1^\omega \phi_1/L\Phi}{1/\Phi + 1}, \tag{25}$$

which shows that λ_1^* increases with Φ. Plugging (25) in (24) yields the equation:

$$\Phi = 1 + \frac{T}{2\beta}\left(\frac{1}{\Phi} + 1\right)\frac{n_1^\omega}{L/\phi_1 + n_1^\omega},$$

which has a unique solution Φ^*. Clearly, $\mathrm{d}\Phi^*/\mathrm{d}T > 0$, so that $\mathrm{d}\lambda_1^*/\mathrm{d}T > 0$ since λ_1^* increases with Φ. In other words, lowering trade barriers triggers the desindustrialization of the small economy. The intuition is easy to grasp. Although foreign competition exerts a downward pressure on domestic prices, hence on wages, in the manufacturing sector, sectorial mobility allows workers to dampen down the wage drop by changing jobs. Lower trade barriers or a larger number of foreign varieties reduce the share of the labor force working in the manufacturing sector by making the service sector relatively more attractive.

As in the foregoing, two types of spatial equilibria may emerge. In the first one, the two cities have the same size, hence $\lambda_{1r}^* = \lambda_1^*/2$ and $\lambda_{2r}^* = \lambda_2^*/2$, when commuting costs are high enough, i.e., $\theta > \delta\theta^\omega$ with

$$\theta^\omega \equiv \frac{(1 + n_1^\omega \phi_1/L)(\phi_1 + 2\phi_2\Phi^*)}{\phi_2^2\Phi^*(1+\Phi^*)}.$$

Since $\mathrm{d}\theta^\omega/\mathrm{d}\Phi^* < 0$ and $\mathrm{d}\Phi^*/\mathrm{d}T > 0$, the condition $\theta > \delta\theta^\omega$ becomes more stringent, and thus urban hierarchy is more likely to emerge when trade costs decrease.

By contrast, when $\theta < \delta\theta^\omega$, the distribution of workers between cities is such that

$$\lambda_A^* - \lambda_B^* = \frac{2\delta\sqrt{\Delta_\omega}}{\theta\phi_1\phi_2} > 0 \qquad \text{with} \quad \Delta_\omega \equiv \phi_1\phi_2^2\frac{(\theta^\omega - \theta)(1 + n_1^\omega\phi_1/L)\Phi^*}{\delta(1+\Phi^*)}.$$

Cumbersome, but standard, calculations show that, once commuting costs take on sufficiently low values, $\lambda_A^* - \lambda_B^*$ increases when T decreases. In this case, more intense competition from the rest of the world implies an employment shift to the larger city where workers consume a wider range of local services.

To sum up, *in a small open economy where commuting costs are not too high, trade liberalization fosters deindustrialization and exacerbates the unequal distribution of jobs between cities.*

5 The Future of Cities

Can NEG models such as those surveyed in the foregoing sections help understand some of the main challenges faced by cities. In what follows, we consider four issues that have important policy implications: (i) The higher number of crimes in larger cities, (ii) the environmental impact of the rapid urbanization in emerging countries like China and India, (iii) the growing share of retirees in Japan and European countries, that is, individuals whose income does not stem from labor, and (iv) the impact of new information and communication technologies, which are often described in the media as a substitute for physical proximity.

5.1 Urban Crime

There is *relatively* more crime in big cities than in small cities (Glaeser and Sacerdote 1999). For example, the rate of violent crime in American cities with more than $250,000$ inhabitants is 346 per $100,000$ inhabitants, whereas in cities with less than $10,000$ inhabitants, the rate of violent crime is 176 per $100,000$. Gaigné and Zénou (2015) provide a microfoundation to these empirical facts by extending the model developed in Section 3.1.

An individual chooses to be either a worker or a criminal. Hence $\lambda_r L = C_r + W_r$ where C_r (resp., W_r) is the mass of criminals (resp., workers) in city r. Individuals are assumed to be heterogeneous in their incentives to commit crimes. More specifically, individuals have different aversions to crime, that is, a higher c means more aversion towards crime. Individuals are rational decision-makers who engage in legal or illegal activities according to the expected utility generated by each activity.

Denoting by ξ the lump-sum amount stolen by a criminal, a worker's income is now given $w_r - \xi C_r$, while the income of a criminal is ξW_r. On average, each worker is "visited" by C_r criminals who each takes ξ; each criminal "robs" W_r workers and takes ξ from each worker. Therefore, the average proceeds from crime are equal to ξL_r. Although there is no direct competition between criminals, they are *indirectly* connected via the total mass of workers living in a city because more criminals induce less workers, which, in turn, lowers pecuniary returns per crime.

Criminals travel less than workers. For simplicity, we normalize criminals' commuting costs to zero. In equilibrium, workers bid away criminals who will live at the city fringe. Urban costs borne by a worker are now given by $\theta W_r/(2\delta)$ instead of $\theta\lambda_r L/(2\delta)$.

A. Criminal Activity and City Size. The indirect utility of a worker is given by $V_r = n_r + w_r^* - \xi C_r - \theta W_r/2\delta + \overline{q}_0$, whereas the indirect utility of a criminal is $V_r^c = n_r + \xi W_r - c + \overline{q}_0$. The individual indifferent between committing a crime and working is such that $V_r = V_r^c$. Assuming that c is uniformly distributed on the interval $[0, 1]$, the fraction of criminals in a city is given by $\psi_r \equiv C_r/\lambda_r$. The equilibrium share of criminals is thus given by the expression:

$$\psi_r^* = \frac{\theta/\delta - 2\left(\mathbf{p}^2/\phi - \xi\right)}{\theta/\delta + 2/\lambda_r L},$$

which is always smaller than 1 when $\mathbf{p}^2/\phi > \xi$, a condition that must hold for a city to exist. Hence the total mass of criminals in the economy depends on the intercity distribution of individuals λ:

$$C(\lambda) = \lambda L\psi_A^* + (1-\lambda)L\psi_B^*.$$

It can readily be shown that $\partial C(\lambda)/\partial\lambda > 0$ for $\lambda \geq 1/2$. Even though wages increases with city size, the agglomeration of activities gives rise a *multiplier effect* through the following two channels. First, a larger city size induces greater expected pecuniary returns because criminals face a larger number of potential victims. Second, a more populated city diminishes the opportunity cost of illegal activity. Even if the amount of stolen wealth by each criminal, that is, $\xi(1-\psi_r)\lambda_r L$, is proportional to city size, *there is proportionally more crime in bigger cities than in smaller cities.*

Furthermore, $\partial\psi_r^*/\partial\theta > 0$ when $\psi_r^* < 1$. Hence, a worse job access leads to more criminal activities within a city. When commuting costs increase, total urban costs also increase, so that the net wages of workers are reduced. This, in turn, leads to a larger fraction of criminals.

B. The Emergence of Large Cities with High Criminal Activity. Individuals choose in which city to reside without knowing their types, which are revealed after their location choices. Therefore, the location of individuals is driven by the inter-city difference in their *expected* utility:

$$\mathbf{EV}_r = \int_0^{\widehat{c}_r} V_r^c \mathrm{d}c + \int_{\widehat{c}_r}^1 V_r \mathrm{d}c = \frac{1}{2}(\psi_r^*)^2 + V_r(\psi_r^*).$$

In addition to the various forces presented in Section 3.1, the location of individuals is affected by the level of criminal activity (Cullen and Levitt 1999). The different mechanisms interact with the decision to become a criminal, hence with the level of wages, urban costs, and λ. The expected utility differential is given by

$$\mathbf{EV}_A - \mathbf{EV}_B = \left(\lambda - \frac{1}{2}\right)\Gamma,$$

where Γ is a positive or negative bundle of parameters. Gaigné and Zénou (2015) show that, when commuting costs take on intermediate values, *the larger city hosts more workers and more criminals.* Furthermore, since $\psi_A > \psi_B$ and $\mathbf{EV}_A = \mathbf{EV}_B$ when $1/2 < \lambda^* < 1$, *workers living in the smaller city are better off than those living in the larger city.*

5.2 Are Compact Cities Ecologically Desirable?

The transport sector is a large and growing emitter of greenhouse gases (GHG). It accounts for 30% of total GHG emissions in the United States and approximately 20% of GHG emissions in the EU-15. Moreover, road-based transport accounts for a very large share of GHG emissions generated by this sector. For example, in the US, nearly 60% of GHG emissions stem from gasoline consumption for private vehicle use, while a share of 20% is attributed to freight trucks, with an increase of 75% from 1990 to 2006. Although new technological solutions will improve energy efficiency, other initiatives are needed, such as mitigation policies based on the reduction of average distances travelled by commodities and people.

A. The Ecological Trade-Off Between Commuting and Transport Costs. We have seen that transporting people and commodities involves economic costs. It also implied ecological costs that obey the fundamental trade-off of Section 2.1: The agglomeration of firms and people in a few large cities minimizes the emissions of GHG stemming from shipping commodities, but increases those generated by longer commuting; dispersing people and firms across numerous small cities has the opposite costs and benefits. If cities are more compact ($\delta \uparrow$), then, keeping population and firms fixed, the costs associated with the former spatial configuration (concentration) fall relative to those associated with the latter (dispersion) because people commute over shorter distances. However, when one recognizes that firms and people choose their location in order to maximize profits and utility, a policy that aims to make cities more compact will affect the intercity pattern of activities by fostering their progressive agglomeration, thus raising the level of GHG within fewer and larger cities. Therefore, the ecological effects of an increasing-density policy are a priori ambiguous (Gaigné et al. 2012).

To illustrate how this trade-off operates, we consider the model of Section 2.1 and assume that the carbon footprint E of the urban system stems from the total distance travelled by commuters within cities (D) and the total quantity of the manufactured good shipped between cities (T):

$$E = e_D D + e_T T,$$

where e_D is the amount of GHG generated by one unit of distance travelled by a consumer, while shipping one unit of the manufactured good between cities generates e_T units of carbon dioxides.

Because consumers are symmetrically distributed on each side of the CBD at the equilibrium, the value of C depends on the intercity distribution of the manufacturing sector and is given by

$$D = \frac{L^2}{4\delta}(\lambda_r^2 + \lambda_s^2).$$

Clearly, the emission of GHG stemming from commuting is minimized when the manufacturing sector is evenly dispersed between two cities ($\lambda_A = \lambda_B = 1/2$).

Regarding the value of T, it is given by the sum of equilibrium trade flows:

$$T = \frac{\phi[4\beta - (4\beta + 1)\tau]}{2(2\beta + 1)\beta}\lambda_A \lambda_B,$$

where $T > 0$ because $\tau < \tau_{\text{trade}}$. As expected, T is minimized when consumers and firms are agglomerated within a single city. Note also that T increases when shipping goods becomes cheaper because there is more intercity trade. Hence, *transportation policies that reduce shipping costs foster a larger emission of GHG.*

Thus, E is described by a concave or convex parabola in λ, so that the emission of GHG is minimized either at $\lambda = 1$ or at $\lambda = 1/2$. Therefore, it is sufficient to evaluate the sign of $E(1; \delta) - E(1/2; \delta)$, which is negative if and only if $\delta > \delta_e$ where

$$\delta_e \equiv \frac{e_D}{e_T} \frac{(2\beta + 1)\beta}{\phi[4\beta - (4\beta + 1)\tau]},$$

which is positive because $\tau < \tau_{\text{trade}}$. Thus, the agglomeration of activities within a single compact city is ecologically desirable if and only if $\delta > \delta_e$. Otherwise, dispersion is the best ecological outcome. Consequently, *agglomeration* *or dispersion is not by itself the more preferable pattern from the ecological viewpoint.* For agglomeration to be ecologically desirable, the population density must be sufficiently high for the average commuting distance to be small enough.

B. Does the Market Yield a Good, or a Bad, Ecological Outcome? As seen in Section 2.1, $\lambda = 1/2$ is a stable equilibrium if the population density δ is smaller than $\delta_m \equiv \phi\theta/\Upsilon(\tau)$. Otherwise, the manufacturing sector is concentrated into a single city. Because δ_m increases with θ and is equal to 0 at $\theta = 0$, while $\delta_e > 0$ is independent of θ, there exists a unique value $\bar{\theta}$ such that $\delta_m = \delta_e$. As a result, the market delivers a configuration that minimizes or maximizes the emissions of GHG. In other words, *the market yields either the best or the worst ecological outcome.*

Consider, first, the case where θ exceeds $\bar{\theta}$. If $\delta < \delta_m$, the market outcome involves two cities. Keeping this configuration unchanged, a more compact city ($\delta \uparrow$) always reduces the emissions of pollutants. Once δ exceeds δ_m, the economy gets agglomerated, thus leading to a downward jump in the GHG emissions. Further increases in δ allow for lower emissions of GHG. Hence, when commuting costs are high, a denser city always yields lower emissions of GHG. Assume now that $\theta < \bar{\theta}$. As in the foregoing, provided that $\delta < \delta_m$, the market outcome involves dispersion while the pollution level decreases when the city gets more compact. When δ crosses δ_m from below, the pollution now displays an upward jump. In other words, when commuting costs are low, *more compact cities are not always ecologically desirable.*

When consumers and firms are mobile, what matters for the total emission of GHG is the mix between city compactness (δ) and city size (λ), thus pointing to the need of coordinating environmental policies at the local and global levels. In other words, environmental policies must focus on the urban system as a whole and not on individual cities. Furthermore, we have seen in Section 2.4 that the internal structure of cities may change with population density. In this case, the ecological effects of an increasing-density policy are even more ambiguous: More compactness favors the centralization of jobs at the city center. Gaigné et al. (2012) point out that, unless commuting to SBDs generates a massive use of private cars, compact and monocentric cities may generate more pollution than polycentric and dispersed cities. By lowering urban costs without reducing the benefits generated by large urban agglomerations, the creation of SBCs may allow large cities to reduce GHG emissions while benefiting from stronger agglomeration economies.

In sum, building tall cities is part of the answer to the question that serves as a title to this section. However, we contend that policy-makers should also pay attention to the structure and number of cities.

5.3 Cities in Aging Nations

The old-age dependency ratio (the ratio people aged 65 and older to people aged 15 to 64) is projected to double by 2050 within the European Union, with four persons of working age for every elderly citizen to only two. This ratio is expected to be lower in the United States, with a rise from 19 to 32%, but higher in Japan, with a rise from 25 in 2000 to 72% in 2050. Such demographic changes are likely to have a major impact on cities because the retirees are driven by location factors that differ from those governing workers' residential choices. Workers' welfare depends on local services, land rent and wages, whereas rentiers' welfare depends only upon local services/amenities and land rent. As a consequence, when the share of old people takes on a sufficiently high value, the process of circular causation sparked by workers' location choice could well be challenged.

To study how the urban system might change as the old-age dependency ratio rises, we consider the model of Section 4.1 in which the population is split between two groups, i.e., the *elderly* and the *workers* whose respective shares are $\rho \geq 0$ and $1-\rho \geq 0$. City A is endowed with an amenity $a > 0$ that is valued only by the elderly, while B is called the working-city. We close the model by assuming that land is collectively owned by the elderly. The income of a retiree is, therefore, given by the aggregate land rent (ALR) divided by the total number of elderly (ρL). Workers and retirees have different unit commuting costs, t and θ, respectively.

Let s_r be the share of elderly people living in city $r = A, B$. City r-population is then given by $\lambda_r = (\lambda_{1r} + \lambda_{2r})(1-\rho)L + s_r \rho L$. Besides λ_{1r} and λ_{2r}, we also have to determine s_r.

Workers and retirees have different unit commuting costs denoted, respectively, by θ and t. Since moving is typically more demanding for the retirees, it is reasonable to assume $t > \theta$. However, the results remain qualitatively the same when $t < \theta$. The inequality $t > \theta$ implies that the elderly live closer to the CBD. Their urban costs UC_r^o are given by

$$UC_r^o = \frac{t}{\delta}\frac{s_r \rho L}{2} + \frac{\theta}{\delta}\frac{(\lambda_{1r}+\lambda_{2r})(1-\rho)L}{2}.$$

whereas workers' urban costs are now as follows:

$$UC_r = \frac{\theta}{\delta}\frac{s_r \rho L}{2} + \frac{\theta}{\delta}\frac{(\lambda_{1r}+\lambda_{2r})(1-\rho)L}{2},$$

which is equal to (5) when $\rho = 0$.

Both workers and retirees choose simultaneously their location according to their respective inter-city utility differential. The asymmetry in the amenity supply implies the following elderly's equilibrium condition $V_A^o - V_B^o = a$ with $V_r^o = n_1 + n_{2r} + ALR/\rho - UC_r^o$. It can be shown that the equilibrium distribution of the elderly between cities is the same when ρL is sufficiently large:

$$s_A^* = \frac{1}{2} + \frac{a\delta}{(t-\theta)\rho L}. \tag{26}$$

Otherwise, $s_A^* = 1$. In both cases, *more than half of the retirees choose to live in the amenity-city.*

When the share of the elderly people is not too high, the economy displays two stable equilibria (Gaigné and Thisse 2009). In the former one, the working-city remains dominant; in the latter one, the amenity-city is the primate city. This could explain why there are contradicting opinions regarding the evolution of urban systems in aging nations. The pattern in which the working-city remains the primate city, while the other city accommodates the larger share of retirees, is the equilibrium that agrees with current empirical evidence (Chen and Rosenthal 2008). The corresponding distribution of workers between sectors and cities is as follows:

$$\lambda_{1B}^* = \frac{1}{4} + \frac{(2\delta - \theta\phi_2)\sqrt{\Delta_a}}{4\theta\phi_1\phi_2} + \frac{\rho}{1-\rho}\left(s_A^* - \frac{1}{2}\right) \qquad \lambda_{2B}^* = \frac{1}{4} + \frac{\sqrt{\Delta_a}}{4\phi_1} \tag{27}$$

where

$$\Delta_a \equiv \phi_1(\phi_1 + 2\phi_2) - \frac{2\theta\phi_1\phi_2^2}{\delta(1-\rho)} < \Delta.$$

The expression (27) shows how the elderly' location choices affect the distribution of activities.

It follows from (26) and (27) that *workers and retirees are not attracted by the same city*. This provides a rationale for recent empirical evidence, which suggests that retirees and workers tend to live separately as the old-age dependency ratio increases. Furthermore, when city A's local government improves its amenity supply ($a \uparrow$), this city attracts a growing number of retirees, which allows the number of jobs in the working-city to rise.

Moreover, an aging population ($\rho \uparrow$) induces the dispersion of services at the expense of the working-city. In other words, an increasing share of retirees may challenge the efficiency of the working-city. As a result, if the agglomeration of manufactures and services generates benefits not taken into account in the model, the economy will incur efficiency losses. However, beyond some threshold the migration of retirees toward the amenity-city raises the level of urban costs and/or decrease the supply of local services in city A. This restores partially the attractiveness of the working-city. Nevertheless, this need not be true for services which still benefit from a big market in the elderly-city. Finally, regardless of old-age dependency ratio, the working-city remains the larger one ($\lambda_B^* > \lambda_A^*$).

To sum up, though in an aging nation the relocation of consumption services weakens the supremacy of working-cities, these ones remain the biggest ones. Indeed, as long as it is more profitable for the bulk of manufactures to congregate, a large share of services is prompted to set up therein. In addition, as the population gets older, cities diverge in their job and demographic structures. Yet, the supply of consumption services in both cities should prevent the complete spatial separation of workers and retirees.

5.4 Do New Information and Communication Technologies Foster the Death of Cities?

Besides transport costs, geographical separation generates another type of spatial friction, that is, *communication costs*. Decreasing communication costs allow the slicing up of the supply chain by reducing the costs of coordinating production across space and time zones. As a result, firms are able to disperse their functions, such as management, finance, R&D, and production, into geographically separated units in order to benefit from the attributes specific to different locations (Aarland et al. 2007). This topic has been largely overlooked in NEG. By contrast, the topic has attracted a lot of attention in the theory of the multinational enterprise (Markusen 2002).

To illustrate the interplay between communication and transport costs, consider a monopoly formed by a *headquarters* (HQ) and one or two production *plants*. The two regional markets can be supplied according to one of the following two organizational forms. First, the firm is *integrated* when its operates a single plant located in the same city as the HQ and ships its output to the other city. This organizational configuration corresponds to the case studied in Section 2. Second, the firm chooses a *horizontal* structure where a plant is set up in each city. In other words, rather than exporting, the firm serves the remote city by establishing a subsidiary in the corresponding market. This implies an additional investment ϕ and a communication cost ζ between the HQ and the plant established in the distant city. In this new setting, agglomeration occurs when the monopolist's activities are spatially integrated while dispersion prevails when the firm operates two spatially separated plants.

The profit of an integrated firm is given by (6), whereas the profit of a horizontal firm is equal to

$$\pi^H = p_A q_A \lambda_A L + (p_B - \zeta) q_B \lambda_B L - \phi w_A - \phi w_B.$$

Consumers' utility is given by (1) where $\gamma = 0$ because the firm is a monopoly. Hence, the inverse demand function in city r becomes $q_r = 1/\beta - p_r/\beta$. The monopoly first chooses its organizational form and, then, the consumer price in each city. The profit-maximizing prices of an integrated firm are given respectively by $p_A^* = 1/2$ and $p_B^* = (1+\tau)/2$. The firm charges prices equal to $p_A^H = 1/2$ and $p_B^H = (1+\zeta)/2$ when it chooses to go horizontal. As expected, higher communication costs increase the price paid by consumers living in city B.

The firm chooses to go horizontal if and only if $\pi^H > \pi$ or, equivalently,

$$\phi w_B < \frac{[(1-\zeta)^2 - (1-\tau)^2]}{4\beta} \lambda_B L.$$

Thus, dispersion is more likely to occur when *communication costs are low and transport costs are high.* In other words, communication and transport costs do not play the same role in firms' spatial organization.

Firms may also choose a *vertical* structure when their headquarters and plants are spatially separated. Since headquarters benefit from a wide range of agglomeration economies, the typical pattern involves the concentration of firms' core functions in large cities, which provide specialized business-to-business services and lie at the source of various types of knowledge spillovers (Davis and Henderson 2008), together with the decentralization of production in plants located in the periphery where land and labor are cheaper. In this case, firms bear communication costs for conducting their production activities at a distance, as well as transport costs for shipping goods to the HQ city.

A firm will never choose to become vertical unless the host-city B is endowed with a comparative advantage relative to the parent-city A. For example, a cost difference creates a motive for the geographical fragmentation of the supply chain (Antras and Yeaple 2014). We assume here that the comparative advantage takes the form of a subsidy $m_B > 0$ paid to the firm. In this case, the firm weighs the cost of offshoring the production of its product against the cost saving from producing at a distant place where production costs are lower. The profit of a vertical firm is as follows:

$$\pi^V = (p_A - \tau - \zeta + m_B)q_A\lambda_A L + (p_B - \zeta + m_B)q_B\lambda_B L - \phi w_B.$$

The profit-maximizing prices are $p_A^V = (1 + \tau + \zeta - m_B)/2$ and $p_B^V = (1 + \zeta - m_B)/2$. The firm chooses to be vertical if and only if $\pi^V > \pi$ or, equivalently,

$$\phi w_B < \frac{(2 + m_B - \tau - \zeta)[m_B - (\lambda_A - \lambda_B)L\tau - \zeta]}{4\beta},$$

with $2 + m_B - \tau - \zeta > 0$, whereas $m_B - (\lambda_A - \lambda_B)L\tau - \zeta$ can be positive or negative. Clearly, the firm will choose to be vertical when transport and communication costs are low relative to the comparative advantage.

This example captures the main trade-offs described in Fujita and Thisse (2006) who have revisited the core-periphery model when firms are free to choose to be integrated or vertical. The core region hosts the headquarters, while wages are lower in the periphery. When communication costs are high, reducing transport costs leads a growing number of firms to be integrated. However, things are different when communication costs are low. For high transport costs, most plants are set up in the core. However, once transport costs are below some threshold, falling communication costs trigger the relocation of activities and the core loses a growing share of plants. This is because the vertical firms are able to benefit from lower production costs in the periphery without losing much business from the core. Eventually, when communication costs have reached a sufficiently low level, the economy ends up with a deindustrialized core, which retains firms' strategic functions. In other words, *decreasing transport costs incite firms to be integrated*, whereas *decreasing communication costs incite firms to get fragmented.* This prediction concurs with Baldwin (2016) who argues that big drops in transport costs and in communication costs are at the origin of two different types of globalization.

Transport costs imply that selling at a distance is more expensive, but these costs do not affect sales on the domestic market. By contrast, producing at a distance imposes communication costs that affect the whole volume of production while shipping the output from the plant to the firm's home market is still required. This is why the spatial fragmentation of the firm needs low communication and transport costs, which is precisely what Fujita and Thisse (2006) have shown.

6 Conclusion

The idea of spatial interaction is central to regional science. Broadly defined, spatial interaction refers to flows across space that are subject to various types of spatial frictions, such as traded goods, migrations, capital movements, interregional grants, remittances, and the interregional transmission of knowledge and business cycle effects. Though the NEG literature has for the most part focused on the mobility of goods and production factors, these issues are at the heart of NEG. Instead of writing one more review of the vast literature produced in the footsteps of Krugman (1991), we have chosen to highlight the role that NEG may play in understanding the process of urban development. More specifically, through several major trade-offs we have covered a range of issues that highlight the working of urban systems. To do so, we have used very simple models, which vastly contrast with the heavy mathematical apparatus employed in the literature.

Cities of the twenty-first century face new and important challenges, such as climate change, aging population, crime, poverty, social exclusion, food security, the supply and management of transportation and communication infrastructure, and competition with a few global cities. It is, therefore, fundamental to have sound theoretical models that can be used as guidelines in developing empirical research and designing new policies. Is NEG a useful tool? We believe that the answer is yes. From the methodological standpoint, NEG has several major merits. First, the decisions made by firms and households are based on land rents, wages and prices, which are themselves endogenous and related to the size and structure of cities. Second, NEG takes into account the fact that households and firms may relocate between and within cities in response to major changes in their economic environment. Last, the various fields of modern economics NEG is connected with and provide a set of tools and concepts that permit to tackle new and challenging issues.

Nevertheless, NEG suffers from a major drawback: It is built on a two-location setting. The new fundamental ingredient that a multi-location setting brings about is that a firm's location is the balance of a system of forces pulling the firm in various directions. More specifically, the *relative position* of a city within the whole network of interactions matters (Behrens et al. 2007; Matsuyama 2017). Another key insight one can derive in a multi-location economy is that any change in the underlying parameters has in general complex impacts that vary in non-trivial ways with the properties of the underlying transport network. When there are only two locations, any change in structural parameters necessarily affects directly either one of the two cities, or both. On the contrary, when there are more than two locations, any change in parameters that directly involves only two cities now generates spatial spillover effects that are unlikely to leave the remaining cities unaffected. More work is called for here but one should not expect a simple answer. Spatial quantitative models, which blend ingredients from NEG and urban economics, could well provide a solution to this quandary (Redding and Rossi-Hansberg 2017). More work is called for, but one should not expect a silver bullet to solve the dimensionality problem.

We conclude with the following remark. So far, the Internet and rapid, safe trips have not proven to be a good substitute for physical proximity in activities where knowledge spillovers, the transfer of information, and trust are critical. At least for now, face-to-face contacts remain the best way to deal with the communication curse. That said, it is hard to predict what place-based constraints will become with the development of new digital technologies.

Nevertheless, it is worth noting with Gaspar and Glaeser (1998) and Goldfarb and Tucker (2019) that the Internet might well be a complement to cities because of the spatial correlation that characterizes many social networks.

References

Aarland, K., J.C. Davis, J.V. Henderson and Y. Ono (2007) Spatial organization of firms: The decision to split production and administration. *RAND Journal of Economics* 38: 480-494.

Albouy D. and G. Ehrlich (2018) Housing productivity and the social cost of land-use restrictions. *Journal of Urban Economics* 107: 101-120.

Albouy, D., G. Ehrlich and M. Shin (2018) Metropolitan land values. *Review of Economics and Statistics* 100: 454-466.

Antràs, P. and S.R. Yeaple (2014) Multinational firms and the structure of international trade. In: G. Gopinath, E. Helpman and K. Rogoff (eds.) *Handbook of International Economics. Volume 4.* Elsevier, New York, pp. 55-130.

Baldwin, R.E. (2016) *The Great Convergence.* Harvard University Press, Cambridge (MA).

Baldwin, R.E., R. Forslid, P. Martin, G. I. P. Ottaviano and F. Robert-Nicoud (2003) *Economic Geography and Public Policy.* Princeton University Press, Princeton.

Baldwin, R.E. and T. Okubo (2006) Heterogeneous firms, agglomeration and economic geography: Spatial selection and sorting. *Journal of Economic Geography* 6: 323-346.

Behrens, K., G. Duranton and F. Robert-Nicoud (2014) Productive cities: Sorting, selection and agglomeration. *Journal of Political Economy* 122: 507-553.

Behrens, K., A.R. Lamorgese, G.I.P. Ottaviano and T. Tabuchi (2007) Changes in transport and non-transport costs: Local vs global impacts in a spatial network. *Regional Science and Urban Economics* 37: 625-648.

Behrens, K. and F. Robert-Nicoud (2015) Agglomeration theory with heterogenous agents. In G. Duranton, J.V. Henderson and W.C. Strange (eds.) *Handbook of Regional and Urban Economics. Volume 5.* Elsevier, New York, pp. 171-245.

Blanchard, O.J. and L.F. Katz (1992) Regional evolutions. *Brookings Papers on Economic Activity* 1: 1-61.

Brueckner, J., J.-F. Thisse and Y. Zenou (1999) Why is Central Paris rich and Downtown Detroit poor? An amenity-based theory. *European Economic Review* 43: 91-107.

Carlino, G. and W.R. Kerr (2015) Agglomeration and innovation. In G. Duranton, J.V. Henderson and W.C. Strange (eds.) *Handbook of Regional and Urban Economics. Volume 5.* Elsevier, New York, pp. 349-404.

Cavailhès, J., C. Gaigné, T. Tabuchi and J.-F. Thisse (2007) Trade and the structure of cities. *Journal of Urban Economics* 62: 383-404.

Chen Y. and S.S. Rosenthal (2008) Local amenities and life-cycle migration: Do people move for jobs or fun?

Journal of Urban Economics 64: 519-537.

Combes, P.-P. and L. Gobillon (2015) The empirics of agglomeration economies. In G. Duranton, J.V. Henderson and W.C. Strange (eds.) *Handbook of Regional and Urban Economics. Volume 5.* Elsevier, New York, pp. 247-348.

Combes, P.-P., G. Duranton and L. Gobillon (2008) Spatial wage disparities: Sorting matters! *Journal of Urban Economics* 63: 723-742.

Cuberes D., J. Roberts and C. Sechel (2019) Household location in English cities. *Regional Science and Urban Economics*, forthcoming.

Cullen, J. and S. Levitt (1999) Crime, urban flight, and the consequences for cities. *Review of Economics and Statistics* 81: 159-169.

Davis, J.C. and J.V. Henderson (2008) The agglomeration of headquarters. *Regional Science and Urban Economics* 38: 445-460.

de Palma, A., Y.Y. Papageorgiou, J.-F. Thisse and P. Ushchev (2019) About the origin of cities. *Journal of Urban Economics*, forthcoming.

Dixit, A.K. and J.E. Stiglitz (1977) Monopolistic competition and optimum product diversity. *American Economic Review* 67: 297-308.

Duranton, G. and D. Puga (2004) Micro-foundations of urban increasing returns: Theory. In J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics. Volume 4.* Elsevier, Amsterdam, pp. 2063-2117.

Fujita, M. (1989) *Urban Economic Theory. Land Use and City Size.* Cambridge University Press, Cambridge.

Fujita, M. and P. Krugman (1995) When is the economy monocentric? von Thünen and Chamberlin unified. *Regional Science and Urban Economics* 25: 505-528.

Fujita, M., P. Krugman and A.J. Venables (1999a) *The Spatial Economy. Cities, Regions and International Trade.* The MIT Press, Cambridge, MA.

Fujita, M., P. Krugman and T. Mori (1999b) On the evolution of hierarchical urban systems. *European Economic Review* 43: 509-551.

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

Fujita, M. and J.-F. Thisse (2013) *Economics of Agglomeration. Cities, Industrial Location and Globalization. Second edition.* Cambridge University Press, Cambridge.

Gaigné, C. and J.-F. Thisse (2009) Aging nations and the future of cities. *Journal of Regional Science* 49: 663-688.

Gaigné, C., S. Riou and J.-F. Thisse (2012) Are compact cities environmentally friendly? *Journal of Urban Economics* 72: 123-136.

Gaigné, C. and Y. Zenou (2015) Agglomeration, city size and crime. *European Economic Review* 80: 62-82.

Gaspar, J. and E.L. Glaeser (1998) Information technology and the future of cities. *Journal of Urban Economics* 43: 136-156.

Glaeser, E.L., J. Gyourko and R.E. Saks (2006) Urban growth and housing supply. *Journal of Economic Geography* 6: 71-89.

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

Glaeser, E.L. and D.C. Maré (2001) Cities and skills. *Journal of Labor Economics* 19: 316-342.

Glaeser, E.L. and B.I. Sacerdote (1999) Why is there more crime in cities. *Journal of Political Economy* 107: 225-258.

Goldfarb, A. and C. Tucker (2019) Digital economics. *Journal of Economic Literature* 57: 3-43.

Gollin, D., R. Jedwab and D. Vollrath (2016) Urbanization with and without industrialization. *Journal of Economic Growth* 21: 35-70.

Handbury, J. and D. Weinstein (2015) Goods prices and availability in cities. *Review of Economic Studies* 82: 258-296.

Helpman, E. (1998) The size of regions. In D. Pines, E. Sadka and I. Zilcha (eds.) *Topics in Public Economics. Theoretical and Applied Analysis.* Cambridge University Press, Cambridge, pp. 33-54.

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

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

Henderson, V. and A. Mitra (1996) New urban landscape: Developers and edge cities. *Regional Science and Urban Economics* 26: 613-643.

Kahn, M.E. and R. Walsh (2015) Cities and the environment. In G. Duranton, J.V. Henderson and W.C. Strange (eds.) *Handbook of Regional and Urban Economics. Volume 5.* Elsevier, New York, pp. 405-465.

Knoll, K., M. Schularick and T.M. Steger (2017). No price like home: Global house prices, 1870-2012. *American Economic Review* 107: 331-353.

Koster, H.R.A. and J. Rouwendal (2017) Historic amenities and housing externalities. Evidence from the Netherlands. *Economic Journal* 127: F396–F420.

Krugman, P. (1980) Scale economies, product differentiation, and the pattern of trade. *American Economic Review* 70: 950-959.

Krugman, P. (1991) Increasing returns and economic geography. *Journal of Political Economy* 99: 483-499.

Markusen, J.R. (2002) *Multinational Firms and the Theory of International Trade.* MIT Press, Cambridge (MA).

Matsuyama, K. (2017) Geographical advantage: Home market effect in a multi-region world. *Research in Economics* 71: 740-758.

Moretti, E. (2012) *The New Geography of Jobs.* Houghton Mifflin Harcourt, Boston.

Okubo, T., P.M. Picard and J.-F. Thisse (2010) The spatial selection of heterogeneous firms. *Journal of International Economics* 82: 230-237.

Ottaviano, G.I.P. and J.-F. Thisse (2004) Agglomeration and economic geography. In J.V. Henderson and J.-F. Thisse (eds.) *Handbook of Regional and Urban Economics. Volume 4.* Elsevier, Amsterdam, pp. 2563–2608.

Ottaviano, G. I. P. and J.-F. Thisse (2005) New economic geography: What about the N? *Environment and Planning A* 37: 1707-1725.

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

Rodríguez-Pose, A. (2018) The revenge of the places that don't matter (and what to do about it). *Cambridge Journal of Regions, Economy and Society* 11: 189-209.

Samuelson, P.A. (1954) The transfer problem and transport cost. II. Analysis of effects of trade impediments. *Economic Journal* 64: 264-289.

Tabuchi, T. (1998) Urban agglomeration and dispersion: A synthesis of Alonso and Krugman. *Journal of Urban Economics* 44: 333-351.

Tabuchi, T., and J.-F. Thisse (2002) Taste heterogeneity, labor mobility and economic geography. *Journal of Development Economics* 69: 155-177.

Tabuchi, T. and J.-F. Thisse (2006) Regional specialization, urban hierarchy, and commuting costs. *International Economic Review* 47: 1295-1317.

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

Tiebout, C.M. (1956) Exports and regional growth. *Journal of Political Economy* 64: 160-164.

Timothy, D. and W.C. Wheaton (2001) Intra-urban wage variation, employment location and commuting times. *Journal of Urban Economics* 50: 338-366.

Vives, X. (1999) *Oligopoly Pricing. Old Ideas and New Tools.* MIT Press, Cambridge (MA).

Wilson, A. (2010) Entropy in urban and regional modelling: Retrospect and prospect. *Geographical Analysis* 42: 364-394.