Frontiers of the New Economic Geography

Masahisa Fujita and Tomoya Mori

August 8, 2005

Abstract

This paper presents an overview of the recent development in the new economic geography (NEG), and discusses possible directions of its future development. Since there already exist several surveys on this topic, we focus on the selected features of the NEG which are important yet have attracted insufficient attention, and also on the recent refinements and extensions of the framework.

Keywords: New economic geography, agglomeration, international trade, economic growth, transport costs

JEL Classifications: R11, R12, R13, R14, F12, F23

Acknowledgement: We are grateful to Kristian Behrens, Yasusada Murata, and three anonymous referees for their constructive comments. This research has been partially supported by The Grant-in-Aid for Research (Nos. 13851002, 16683001) of Ministry of Education, Science and Culture in Japan, and by the Grant-in-Aid for the 21st Century COE "Interfaces for Advanced Economic Analysis".

*Institute of Economic Research, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501 Japan. Phone: +81-75-753-7122. E-mail: fujita@kier.kyoto-u.ac.jp.
†Institute of Developing Economies/JETRO, 3-2-2 Wakaba, Mihama-ku, Chiba, 261-8545 Japan.
‡Institute of Economic Research, Kyoto University, Yoshida-honmachi, Sakyo-ku, Kyoto, 606-8501 Japan. Phone: +81-75-753-7121. E-mail: mori@kier.kyoto-u.ac.jp.
1 Introduction

This paper presents an overview of recent development in the new economic geography (NEG), and discusses possible directions of its future development. As well known now, the NEG represents a new branch of spatial economics, which aims to explain the formation of a large variety of economic agglomeration in geographical space, using a general equilibrium framework.

Agglomeration or the clustering of economic activities occurs at many geographical levels, having a variety of compositions. At one extreme lies the core-periphery structure at the global scale. In 2000, for example, the NAFTA yielded 35% of the world GDP, EU (15 countries) 25%, and East Asia 23%: thus, 83% of the world GDP concentrated in the three regions. In 1980, the corresponding shares were 27% for NAFTA, 29% for EU, and 14% for East Asia; three regions together 70%. Hence, the concentration of the world GDP in the three regions have been intensifying recently. Furthermore, Hall and Jones [74] observe that high-income nations are clustered in small cores in the Northern Hemisphere and that productivity per capita steadily declines with distance from these cores. Strong regional disparities within the same country imply the existence of agglomerations at another spatial scale. In France, for example, the Île-de-France (the metropolitan area of Paris), which accounts for 2.2% of the area of the country and 18.9% of its population, produces 30% of its GDP.

Regional agglomeration is also reflected in large varieties of cities, as shown by the stability of the urban hierarchy within most countries (Eaton and Eckstein [50], Dobkins and Ioannides [44], Mori, Nishikimi and Smith [124]). Cities themselves may be specialized in a small number of industries, as are many medium-size American cities (Henderson [81]). However, large metropolises like New York and Tokyo are highly diversified in that they nest many industries that are not related through direct linkages (Chinitz [36], Fujita and Tabuchi [67]). Industrial districts involving firms with strong technological or informational linkages, or both (e.g., the Silicon Valley or the Italian districts engaged in more traditional activities) as well as factory towns (e.g., Toyota city in Japan and Hershey in US) manifest various types of local specialization. Therefore, it appears that highly diverse size and activity arrangements exist at the regional and urban levels.

At a very detailed extreme of the spectrum, agglomeration arises in the form of large commercial districts set up in the inner city itself (think of Soho in London, Montparnasse in Paris, or Ginza in Tokyo). At the lowest level, restaurants, movie theaters, or shops selling similar products are clustered within the same neighborhood, or the clustering may take the form of a large shopping mall.

It is also important to notice that all these different types of agglomerations at different levels are embedded in a larger economy, altogether forming a

---

1 See Combes, Duranton and Gobillon [38] for a recent evidence for the wage disparities across French regions.
2 The existing NEG models basically focus on the formation of agglomerations at the city level or at a larger regional scale, rather than the formation of small-scale agglomerations such as clustering of restaurants and shops within a city.
complex system. Understanding all such phenomena is critical for the design of effective urban and regional policies.

To some extent, economic activities are spatially concentrated because of dissimilarities in such natural features as rivers, harbors and mineral deposits, or *first nature*. However, the impact of the *first nature* on the spatial distribution of economic activities is not difficult to explain within the traditional economic theory based on competitive paradigm. Thus, the focus of this paper is on the recent contributions to economic modeling of endogenous mechanisms (or *second nature*) leading to agglomeration. Furthermore, an economic model of agglomeration should explain both concentration and dispersion: in France, for example, why so many people live in Île-de-France and also why so many people do not; or, in New York Metropolitan Area, why so many people work in Manhattan and also why so many other people do not (Krugman[102]). To put it in another way, an economic model of agglomeration is expected to provide a general equilibrium story about the centripetal forces that pull economic activities together and the centrifugal forces that push them apart, relying on the trade-off between various forms of increasing returns and different types of mobility costs.

The NEG is an analytical framework initiated by Paul Krugman [96][97] in early 1990s in order to explain the formation of a large variety of such economic agglomerations in geographical space, and has grown as one of the major branches of the spatial economics today. To date, the NEG remains to be the only general equilibrium framework in which the location of agglomerations is determined explicitly through a microfounded mechanism.

Like any other branch of science, the NEG has precursors for this purpose. Since the 1970s, by incorporating *nonmarket interaction* (externalities) several urban economic models were developed for endogenous formation of the central business district (CBD) within a city. Fujita and Ogawa [65] further explored in this direction, and showed the possibility of the formation of multiple business districts within a city. While such intra-urban agglomeration is one of the possible application of the NEG, it has put more emphasis on the agglomeration within a larger geographical space such as in the system of regions within a country as well as across countries. Theory of urban systems initiated by Henderson [79][80] remains as the workhorse approach for research into the actual distribution of sizes and types of cities. Also, due to its simple model structure, it still continues to be the most popular model on which micro-foundation for the agglomeration economies are embedded. However, the urban-system models a la Henderson cannot be used to explain the location, and hence, the spatial

---

3 For extensive review of the historical development of the economics of agglomeration, see Fujita[56], Fujita and Thisse [68], and Duranton and Puga[46].

4 To the best of our knowledge Solow and Vickrey[150] was the first economic model in which the land-use pattern in an urban area is determined without a priori assumption of any center. The endogenous CBD formation based on nonmarket interactions dates back to Beckmann[10]. See also for Fujita and Smith [66] for a survey of further development in this direction during 1980s.

5 There also exist several urban-system models incorporating micro-founded mechanisms such as matching externalities. For a survey, see Duranton and Puga [46].
patterns of agglomeration given that intercity spatial structure is not explicitly considered in these models.\footnote{There is one model of the Henderson-type city systems by Abdel-Rahman\cite{1} that explicitly considers transport costs between cities, although the city location is not endogenous in this model.}

In this paper, we discuss the development and important features of the NEG. Since there already exist several surveys on this topic,\footnote{For the historical development, see Fujita and Krugman\cite{59}. For the basic models and their implications, see Ottaviano and Puga\cite{136}, Duranton and Puga\cite{46, Sec.2.2}, Ottaviano and Thiss\cite{138, Sec.3.4}, and Fujita\cite{56}. See also Henderson\cite{82} for a different interpretation of the NEG. For the impact of transport costs on the agglomeration/dispersion patterns, see Fujita and Mori\cite{64}.} we focus on the selected features of the NEG which are important yet have attracted insufficient attention, and also on the recent refinements and extensions of the framework.

The rest of the paper is organized as follows. In Section 2, the basic elements of the NEG are described, and key results of the (first-generation models of the) NEG are summarized. In particular, to complement existing surveys, we put special emphasis on the fact that the NEG is the first general equilibrium framework in which the spatial distribution of agglomerations (i.e., number, size and location of agglomerations) can be explained through a microfounded mechanism. In Section 3, we review recent refinements and extensions of the original framework of the NEG. In Section 4, some related empirical works are introduced. Finally, we discuss future research directions in Section 5.

2 The new economic geography: first-generation models

There are four key terms for the (first-generation) NEG. The first is the general equilibrium modelling of an entire spatial economy which sets apart this approach from that of traditional location theory and economic geography. The second is increasing returns or indivisibilities at the level of individual producer or plant, which is essential for the economy not to degenerate into “backyard capitalism” (in which each household or small group produces most items for itself). Increasing returns in turn lead to the market structure characterized by imperfect competition. The third is transport costs (broadly defined), which makes location matter. Finally, the locational movement of productive factors and consumers is a prerequisite for agglomeration.

There are three classes of models in the NEG: core-periphery models, regional and urban system models, and international models. We briefly describe each below.

2.1 Core-periphery model

The core-periphery model, introduced in Krugman \cite{96} provides a basic introductory framework for the NEG. It illustrates how the interactions among
increasing returns at the level of the firm, transport costs and factor mobility can cause spatial economic structure to emerge and change.

There are two regions, two production sectors (agriculture and manufacturing), and two types of labor (farmers and workers). The manufacturing sector produces a continuum of varieties of a horizontally differentiated product; each variety is produced by a separate firm with scale economies, using workers as the only input. The agriculture sector produces a homogeneous good under constant returns, using farmers as the only input. Workers are freely mobile between regions; whereas farmers are immobile, distributed equally between the two regions. Finally, the trade of manufactures involves a positive transport cost (in an iceberg form).

In this model, the immobility of farmers is a centrifugal force because they consume both types of goods. The centripetal force is more complex, involving a circular causation. First, if a larger number of firms locate in a region, a greater number of varieties are produced there. Then, workers (who are consumers) in that region has a better access to a greater number of varieties in comparison with workers in the other region. Thus, (other things being equal) workers in that region get a higher real income, inducing more workers to migrate towards this region. Secondly, the resulting increase in the number of workers creates a larger market than the other region, which in turn yields the home market effect (HME) familiar in international trade (Krugman[94]). That is because of scale economies, there is an incentive to concentrate the production of each variety in only one region; because of the transport cost, (other things being equal) it is more profitable to produce in the region that offers a larger market, and ship to the other. This implies the availability of even more varieties of differentiated goods in the region in question. In short, the centripetal force is generated through a circular causation of forward linkages (the incentive of workers to be close to the producers of consumer goods) and backward linkages (the incentive for producers to concentrate where the market is larger).

If forward and backward linkages are strong enough to overcome the centrifugal force generated by immobile farmers, the economy will end up with a core-periphery pattern in which all manufacturing is concentrated in one region. The core-periphery pattern is likely to occur (i) when the transport cost of the manufactures is low enough, (ii) when varieties are sufficiently differentiated, or (iii) when the expenditure on manufactures is large enough.

Agglomeration need not occur, of course. However, a small change in critical parameters can “tip” the economy, from one in which two regions are symmetric and equal to one in which tiny initial advantages cumulate, turning one region into an industrial core and the other into a deindustrialized periphery. That is, the dynamics of the model economy are subject to catastrophic bifurcations, points at which their qualitative character suddenly changes.

### 2.2 Urban and regional systems

Two-location stories are helpful builders of intuition; yet empirical economic geography must cope with a world in which activities are spread over continuous
space. In the urban and regional system version of the NEG, the most popular modelling strategy is to focus on the spatial distribution of agglomerations (i.e., number, size, spacing, and inter-industry spatial coordination of agglomerations) while abstracting from the internal spatial structure of agglomeration (i.e., a city in these NEG models is represented by a point in the location space). The first such attempt was the “race-track economy” model by Krugman [98] which is a straightforward extension of the Krugman’s core-periphery model to the twelve regions around the circumference of a circle, like a clock, and goods must be transported along the circumference. In this model, starting with any initial distribution of economic activities which was nearly uniform across the twelve regions, the simulation always ended up with all manufacturing equally agglomerated into just two regions which located exactly in the opposite side, leading to the self-organization of a central-place system. This surprising result was later proven analytically (in a more general setting with a continuum of locations) by Krugman [101] using Turing’s [157] approach for morphogenesis in biology.

An alternative, perhaps more realistic approach has been proposed by Fujita and Krugman [58]. In the series of papers based on this approach, the location space is given by the real line along which land is distributed uniformly. All workers in the economy are now assumed to be identical and free to choose their location and occupation. The agricultural good is produced now using both land and labor. Finally, transport costs are assumed to be positive for both the agricultural and industrial goods. In this model, only the agricultural land is the immobile factor, which is the source of the centrifugal force. The approach starts with a von Thünen’s “isolated state”: a city, defined as a concentration of manufacturing, surrounded by an agricultural hinterland. (the NEG formulation makes it possible to have this state a fully defined equilibrium, in which the existence of the central city is derived from the effects of forward and backward linkages, rather than simply being assumed.) Then one gradually increases the population of the economy as a whole. Eventually the outer reaches of the hinterland become sufficiently far from the center that it becomes worthwhile for some firms to “defect,” giving rise to a new city; further population growth gives rise to still more cities; and so on.

Key to this approach is the recognition that the attractiveness of any given location for manufacturing can be represented by an index of “market potential” derived from the underlying economics (Krugman [98]; but the idea of market potential goes back to Harris [76], and this new work can be regarded as justification of that approach). The process of change in the economy can then be regarded as involving a sort of coevolution in which market potential determines where economic activity locates, and the shifting location of that activity in turn redraws the map of market potential. In particular, the market potential of a given industry sharply decreases as moving away from a city in which this industry is agglomerated, then start increasing once again after a certain (industry specific) distance, which gives the microfoundation for the notion of agglomeration (or urban) shadow (see Fujita and Krugman [58]) As noted, the (geographical) size of the shadow is specific to industry, it is larger
if the industry provides more differentiated goods, and/or if the transport costs for these goods are lower.

The city-evolution approach ends up suggesting that despite the existence of many possible equilibria, there should be some predictable regularities in spatial structure. Once the number of cities has become sufficiently large, the size of and distance between cities tend to settle down at a roughly constant level determined by the relative strength of centripetal and centrifugal forces (Fujita and Mori[63]), providing some justification for the central place theory of Lösch [107]. If there are multiple industries that differ in terms of scale economies and/or transport costs, the economy tends to develop a hierarchical structure (Fujita, Krugman and Mori[60]) reminiscent of Christaller [37]. Furthermore, the falling transport costs for manufactures relatively to those for agriculture may eventually lead to the formation of a megalopolis consisting of large core cities that are connected by an industrial belt, i.e., a continuum of small manufacturing cities (Mori[121]). The process of megalopolis formation is characterized by successive filling-in’s of new small cities between large (old) cities leading to the formation of an industrial belt, which has close resemblance to the process observed in the actual formation of megalopolis along the East coast of the US as reported by an economic geographer, Gottman[71].8 So this line of work provides a link back to some of the older traditions in location theory and economic geography.

And there is one other payoff to such an evolutionary modeling: it offers an interesting viewpoint on the role of natural geography in determining economic geography. Anyone who looks even casually at the real geography of economic activity is struck by the important degree of arbitrariness or, at best, historicity involved: New York is New York because of a canal that has not been economically important for 150 years; Silicon Valley, as we know it, exists because of the vision of one Stanford official two generations ago. Yet rivers and ports surely do matter. In fact, in new geography models in which a system of cities evolves, these observations are in effect reconciled. Favorable aspects of a location, such as availability of a good harbor, typically have a “catalytic” role: they make it likely that, when a new center emerges, it will be there rather than some other location in the general vicinity. But once a new center has become established, it grows through a process of self-reinforcement, and may thus attain a scale at which the initial advantages of the location become unimportant compared with the self-sustaining advantages of the agglomeration itself. In an odd way, natural geography can matter so much precisely because of the self-organizing character of the spatial economy.

2.3 Agglomeration and trade

In the previous two types of models, namely, core-periphery, urban and regional system, factor mobility has been a key element in creating agglomeration. But

---

8 Notice that in the two-region model, it is impossible to distinguish between this formation of a megalopolis and formation of many small “discrete” cities as in Fujita and Mori [63]. They both appear as the same dispersion (into the two regions).
in practice the concentration of production is greater than that of resources, in the sense that not every agglomeration is an important producer in every industry. There are many cities specialized in a narrow range of industries such as Detroit and Hollywood. Can new economic geography-type models shed light on such industrial concentration, or must one appeal to other forces not present in the basic approach?

The answer is that while more diffuse, hard-to-model forces like informal diffusion of information surely play an important role in creating and sustaining real-world industrial concentrations, it is also possible with a small modification of the core-periphery approach to shift the focus from agglomeration of resources to geographical concentration of particular industries. Such a shift of focus is especially in the analysis of international specialization and trade, defined in our case as models in which labor is immobile among locations.

The key is to allow for a vertical structure of production in which one or more upstream sectors produce inputs for one or more downstream sectors while both upstream and downstream producers are subject to increasing returns and transport costs. As Venables[159] showed, this immediately means that there are backward and forward linkages that tend to concentrate the upstream and downstream producers in a single location. That is, producers of intermediate goods have an incentive to locate where they have the largest market, which is where the downstream industry is; and producers of final goods have an incentive to locate where their suppliers are, which is where the upstream industry is located.

To simplify, Krugman and Venables[105] assumed that the upstream and downstream industries are really the same—that is, that the same goods are consumed and used as inputs to production of other goods. This leads to a formal model of industry concentration that is algebraically isomorphic to the core-periphery model, with only a slight reinterpretation of the meanings of the symbols.

Alternatively, one can assume a more realistic input-output structure in which each upstream industry provides inputs to several downstream sectors, and conversely. Then, it becomes possible to identify the characteristics of the input-output matrix which leads to the formation of industrial clusters, and also about the sequence in which regions will industrialize as world markets expand (Puga and Venables[143]).

Krugman and Venables[105] show how a gradual process of growing world trade due to falling transport costs can first cause the world to divide spontaneously and arbitrarily into a high-wage, industrialized “North” and a low-wage, primary-producing “South”; then, at a later date, cause the South to rise again at the North’s expense.

Finally, while the models above concern agglomeration/specialization across countries, Krugman and Livas Elizondo[104] provided a way to study the impact of external trade on internal geography. In particular, they have suggested

---

9See also Behrens et. al [22] for a more recent attempt using alternative framework (reviewed in Section 3.1).
that an increase in the access to foreign markets might weaken core-periphery
patterns within developing countries. Tomiura[156] also found evidence that
increasing import penetration weakened industrial concentration within Japan.

3 Frontiers in the theory

The original NEG models rely heavily on the specific functional forms on utility
and production functions, transport technology and so on. The obvious next
step is to work with an alternative set of functional forms and technological
assumptions, and investigate the robustness of the results. Then, the framework
of the NEG itself must be extended further. In this section, we review recent
refinements and extensions of the original NEG setup.

3.1 Monopolistic competition

The original NEG models adopted the Dixit-Stiglitz monopolistic competition
model, which in association with iceberg transport costs yielded fairly tractable
general equilibrium model for economic agglomeration. Under the Dixit-Stiglitz
monopolistic competition model with iceberg transport costs, however, the own-
price elasticities of demands are constant, identical to the elasticities of substi-
tutions, and symmetric across all varieties. As a result, the equilibrium mark-up
of each firm is independent of the spatial distribution of firms and consumers.
Moreover, since the iceberg assumption implies that the transport costs are
multiplicative to the f.o.b. price of the product, any increase in the price of the
shipped good is associated with a proportional increase in its transport costs,
which in many cases unrealistic.

To overcome this shortcoming, Ottaviano, Tabuchi and Thisse (OTT) [137]
proposed a monopolistic competition model using quasi-linear utility with quadratic
sub-utility and linear (additive) transport costs. The OTT model successfully
incorporated the pro-competitive effect, i.e., profit-maximizing prices are decreasing
in the mass of competing firms. The additive transport costs are also useful
in analyzing the relationship between the spatial pricing policies and agglomer-
ation (Ottaviano [135, Ch.3]).

Focusing on the role of pro-competitive effects and additive transport costs
modelled in the OTT framework, there have been vigorous theoretical develop-
ments in the recent years. In particular, in this context, unlike the original
NEG, the interregional trade of differentiated goods does not always take place
in two ways. Since the differentiated varieties are imperfect substitutes, firms
selling their output outside their own region face two opposing constraints: (i)
they must set a price that is high enough to cover trade costs; (ii) they must set

\footnote{It should be noted that the same utility function has previously been introduced by Vives[?]. Thus, the true contribution of the OTT is its extension to incorporate additive transport costs, and to build the model in a general equilibrium setup.}

\footnote{Note also that under this pro-competitive effect, lower transport costs for products implies lower price of these products, while the f.o.b. prices were independent of transport costs under the Dixit-Stiglitz-iceberg framework.}
a price that is low enough relative to the local price. Thus, if the interregional
distribution of firms is asymmetric, it is more likely that firms in the larger
region can profitably export to the high margin periphery (where price compe-
tition is mild), whereas firms in the small region cannot profitably export to
the low margin core (where price competition is fierce). Behrens[17] has shown
that in fact the autarky is an endogenous outcome when transport costs are
sufficiently high, while Behrens[19] found that a sufficiently asymmetric initial
industry distribution under autarky leads to asymmetric patterns of trade given
transport costs decrease, where differentiated goods are traded only one way.12

The quasi-linear specification of the OTT model, however, gives their frame-
work a partial equilibrium flavor, and in particular, income effect is absent in
their utility function. More recently, Behrens and Murata [24], using the ad-
ditively quasi-separable function, have developed a general equilibrium model
of monopolistic competition yielding both income and pro-competitive effects.
Their profit-maximizing price has the following properties: (i) increasing in mar-
ginal costs of production; (ii) decreasing in the number of competing firms; (iii)
increasing in consumers’ expenditure. While its application to the spatial econ-
omy is still in progress, they have shown the pro-competitive effects have major
influence on the well-known results in international trade, economic develop-
ment and growth under the Dixit-Stiglitz models. More extensive applications
of this new alternative formulation of monopolistic competition in the context
of economic geography are awaited.

3.2 Homogeneity of workers

All the NEG models reviewed so far assume homogeneity of (mobile) workers.
Murata[130] and Tabuchi and Thisse [153] introduced taste heterogeneity in res-
dential location as another source of dispersion forces in the NEG framework.
Replacing replicator dynamics by probabilistic migration dynamics, both mod-
els captured the idea that when consumers have heterogeneous attachments to
(non-market) local characteristics, they do not necessarily react in the same way
to the difference in the (market-mediated) real wage across regions. Unlike the
standard core-periphery model, this heterogeneity generates partial agglomera-
tion in equilibrium. This suggests that in order to fully understand the spatial
distribution of economic activities, we must take into account both market and
non-market factors.13

12 For further developments using the OTT model, see Behrens[18] for new results on the
home-market effect, and Behrens, Gaigne, Ottaviano and Thisse[21] on the impacts of inter-
national and interregional transport costs on the interregional population distribution, the
terms of trade, and welfare.

13 In both models, however, the interpretation of the stability of equilibria is rather unre-
alistic. Since consumers’ preference for location is determined randomly at each moment, it
is not only that consumers are heterogeneous in their preference for residential location, but
also this preference keeps randomly changing. It is different from the more plausible tradi-
tional formulation of taste heterogeneity in residential location. For instance, Mansoorian
and Myers[113] and Sakashita[148] considered “attachment to home,” i.e., other things being
Mori and Turrini[126] introduced heterogeneity in the innate skill levels among workers. Products are both horizontally and vertically differentiated, and producing higher quality goods requires workers with higher skills. Selling to customers based in a different location entails iceberg-type transport costs and additional communication costs consisting of fixed loss of quality (i.e., non-iceberg transport costs). In this setup, it has been shown that the presence of pecuniary externalities creates a mechanism which always promotes spatial sorting of workers according to their skill levels. In particular, given even a marginal presence of skill heterogeneity and non-iceberg transport costs in the original core-periphery model by Krugman, in all stable equilibria, workers sort themselves across regions in terms of skill level where workers with higher skill choose to stay in the location where aggregate skill and income is higher, while the less skilled stay in the other.

Amiti and Pissarides[5] incorporated firm-worker matching externalities by introducing worker heterogeneity a la Helseley and Strange[?]. As in the standard NEG models, there are constant-returns “agricultural” sector, and monopolistically competitive “manufacturing” sector, both of which uses labor as sole input. For the manufacturing, however, the effective labor depends on the match quality. In their two-country setup, firms can move across countries while workers cannot. In this context, the matching externalities are the reason for the agglomeration of firms in one country, while the proximity to the demand, i.e., immobile workers, tends to disperse firms between the two countries.

3.3 Homogeneity of location space

In the most NEG models, location space is assumed to be homogeneous, hence the locations of agglomerations are determined purely by the second nature: history and cumulative process of forward and backward linkages. However, there are a few interesting situations in which the first nature advantage of location plays a non-trivial role through its interaction with the second nature advantage.

Matsuyama and Takahashi[117] have investigated interaction between agglomeration economies and regional comparative advantage by developing a two-region NEG model with non-tradable differentiated goods. These regions are heterogeneous, since each has comparative advantage in different good. Thus, there is a cost of core-periphery structure since the formation of core means high production cost for the goods in which the periphery has comparative advantage.

Krugman [99][100] have modelled the interaction between geographical centrality as the first nature advantage and the agglomeration economies. Fujita and Mori[62] investigated the role of natural ports (transport hub) as a determinant of industrial agglomeration. In the models with a continuous location space, such as Krugman[99] and Fujita and Mori[62], the “geographical centrality effect” and “hub effect”, the impact of the presence of advantageous location, equal, each individual prefers to live in his home town, and of course, this home town remains the same forever.
appears in the market potential function, and generates a local peak of market potential around at this location.

Martin and Rogers[?], in the two-country setup, studied the situation in which domestic and international transport infrastructures are asymmetric between the countries. This asymmetry is incorporated by decomposing the iceberg-transport costs into domestic and international components. By modelling financing schemes for these infrastructures, a number of interesting policy implications are drawn on the relationship between industrial location and the quality of domestic and international transport infrastructure.

In the OTT framework, Ago, Isono and Tabuchi[3] have shown that in the presence of pro-competitive effect, the central location is more likely to be disadvantageous due to the proximity to a larger number of competitors. Behrens, Lamorgese, Ottaviano and Tabuchi [22] developed a multi-country NEG model in which “economic distance” (e.g., tariffs, transport costs) between countries could be asymmetric, and have derived a number of intriguing implications regarding the impacts of the improvements of transport infrastructure and expenditure shifting.

3.4 Multi-unit firms and spatial fragmentation

All the NEG models reviewed so far assume that firms are spatially integrated, with each firm conducting its entire operation at a single location. In reality, however, a growing number of firms choose to break down their production process into various stages spread across different regions or countries. This spatial fragmentation of production aims at taking advantage of differences in technologies, factor endowments, or factor prices across places.

To investigate the possible economic consequences of the process of international fragmentation, Fujita and Thisse [70] introduced a general equilibrium model of monopolistic competition with two countries, in which each firm has two units, a headquarter (HQ) and a manufacturing plant. HQs use skilled labor, whereas plants use headquarter services together with unskilled labor. The HQ and plant of each firm need not to be located in the same country. The paper focuses on two distinct facts of globalization: the decrease in the trade costs of goods and the decline of communication costs between HQs and plants within firms. It is shown, in particular, that the decline of communication costs may eventually triggers the re-location of plants into the periphery.

Fujita and Gokan [57] further extended the spatial-fragmentation model of Fujita and Thisse [70] by introducing the possibility of multiple plants for each product, which are to be built in separate countries. It is shown that, with decreasing communication costs, firms producing low trade-cost goods (such as electronics products) tend to concentrate their plants in low wage countries, whereas firms producing high trade-costs (such as automobiles) tend to have multiple plants serving to segmented markets.

Given that the spatial fragmentation of production represents one of the main ingredients of the economic process of globalization, this line of research
deserves further efforts.\textsuperscript{14}

3.5 Agglomeration and growth

The first-generation models of the NEG are essentially static: once the economy reaches an equilibrium, no further change occurs in the economy unless parameters are exogenously varied. In other words, first-generation models do not account for the possible impact of agglomeration on the rate of innovation, which in turn is likely to influence further the geographical distribution of economic activities and welfare. It is, therefore, essential to extend the NEG framework into a dynamic setting.

Clearly, space and time are intrinsically mixed in the process of economic development, but the study of their interaction is a hard task. Indeed, because both agglomeration and growth are complex phenomena in themselves, one should expect any integrated analysis to face many conceptual and analytical hurdles. Not surprisingly, therefore, the field is still in its infancy and relevant contributions are not many. Yet, the task is not out of reach. Because both the "new" theories of growth and "new" economic geography share the same basic framework of monopolistic competition, there exists a solid foundation for cross-fertilization between the two fields. The existing contributions that have recently explored the mutual influences between growth and location exploit this formal analogy—see e.g. Waltz[160], Baldwin[11], Martin and Ottaviano[114][115], Baldwin et al. [15], Yamamoto[161]. In particular, Baldwin et al. [16] present several models of growth and agglomeration which are analytically tractable, and examines the welfare impact of various public policies.

Most existing models of growth and agglomeration adopt international settings where the migration of workers across regions or countries are not permitted. This is because the introduction of workers’ migration into an endogenous growth model under perfect foresight raises unsuspected problems that are discussed below. Despite those difficulties, it is possible to derive some tentative conclusions that appear to be reasonable. To the best of our knowledge, to date there exist only four papers that allow for labor mobility in a multi-regional (or multi-city) endogenous growth model under perfect foresight: Waltz [160], Baldwin and Forslid [12], Black and Henderson [27] and Fujita and Thisse[69].

To illustrate the typical approach in these models, we briefly summarize below the work by Fujita and Thisse[69].

It represents a simple model of endogenous growth for a two-region economy, which combines a Krugman-type core-periphery model with a Grossman-Helpman-Romer-type model of endogenous growth, with horizontally differentiated products. Specifically, the core-periphery model is put in a dynamic framework by adding a research and development sector that uses skilled workers to create new varieties for the modern sector, so that the number of varieties

\textsuperscript{14}See Duranton and Puga[47] for a non-NEG formulation of the location problem of multi-unit firms.
produced in the economy growth with time, while forward-looking behavior and migration are formalized in the same spirit as in Ottaviano, Tabuchi and Thisse[137]. The innovation activity in the R&D sector involves knowledge externalities among skilled workers, which occur more intensely in the same region than across the regions. Thus a more agglomeration of skilled workers in a region leads to a higher productivity of skilled workers in the R&D sector in that region. This leads to the possibility that the additional growth spurred by agglomeration may yield a Pareto-dominant outcome. That is, when the economy moves from dispersion to agglomeration, innovation proceeds at a faster pace. As a consequence, even those who stay put in the periphery are better off than under dispersion, provided that the growth effect triggered by the agglomeration is strong enough.

It must be noted, however, that all the existing studies so far skirt the potentially interesting possibility of the cyclical cross-regional-migration of skilled workers in the economy. This possibility for cyclical change in places arises from the forward-looking behavior of workers under the presence of saving opportunities for averaging the lifetime consumption expenditure to take the advantage of regional differences in wage rates and price indices. Although the existing studies exclude such a possibility by assumption, the phenomenon of the cyclical migration of skilled workers have an interesting and important implication on the skill / knowledge transfer among regions, as discussed in Section 5.4 below.

3.6 Application to economic development

Most NEG models assume homothetic preferences over agricultural and manufactured goods. This in turn means a constant expenditure share of each final good. This, together with an exogenous employment share as in the standard core-periphery model, may pause a limitation in the application to the problem of economic development. Murata [129][131] explored the role of nonhomothetic preferences in urbanization/agglomeration. He showed how the two laws in economics — the Engel’s[52] law of the demand shift from agriculture to manufacturing and the Petty’s[139] law of the labor reallocation from the primary industry to the secondary or tertiary industries — can be reconciled with an evolution of a spatial economy.15

Yamamoto[162] proposed a two-region model in which manufactured goods can be produced either with traditional constant-returns technology or with modern technology using differentiated intermediate goods. He have shown that the industrialization, i.e., the shift from traditional to modern technology in the manufacturing production, takes place when transport costs fall sufficiently so that the markets in the two regions are sufficiently integrated.

15 Three other models of the NEG incorporated nonhomothetic preferences. Puga and Venables[143][144] introduced nonhomothetic preferences in an international model without internal geography. Ottaviano, Tabuchi, and Thiss [137] also used nonhomothetic preferences. However their quasi-linear specification implies a decreasing (rather than increasing) expenditure share of manufactured good according to a rise in the wage rate, while the defining characteristics of economic development are often summarized as a decreasing expenditure share of food and a decreasing employment share of agriculture.
4 Frontiers in empirical studies

There are many empirical studies related to the NEG. But, very few of them are NEG specific. Here, we focus on the literature that are specific to the NEG.

4.1 Market potential and the spatial wage structure

Hanson[75] examined the spatial correlation between the market potential and the wages across counties in the US. Two specifications of market potential were considered, where one is based on the traditional Harris[76], i.e., the market potential is higher given a greater proximity to demand (income), while the other is based on Helpman[78] where the centrifugal force in Krugman’s[96] is replaced by land for housing.16 In both specifications, he found significant positive correlations. Brakman et. al[30] applied Hanson’s[75] model to the case of Germany, and also found positive results.17 A similar exercise was conducted by Redding and Venables[146] in the cross-country context. They also found evidence that the proximity to markets and sources of supply is significant and quantitatively important in explaining cross-country variation in per capita income even after controlling for various economic, geographic, social and institutional characteristics.

Redding and Sturm[145] looked at the German division between 1945 and 1989 as a natural experiment to provide evidence for importance of market access for regional economic development. They concluded that the decline of the border cities can be entirely accounted for by their loss in market access by this division, and is neither driven by differences in industrial structure nor differences in the degree of war related destruction.

Recent work by Combes, Duranton and Gobillon[38] decomposed the source of spatial wage disparities into heterogeneity of workers’ characteristic (which they interpreted as the difference in skill level), local endowments, and net local interactions effects (e.g., employment density, industrial diversity, market potential) using the large panel data of French workers, and with careful controls for the endogeneity.18 They concluded that the major source of the spatial wage disparity is the heterogeneity of workers (which they interpret as skill) explaining about half of the variation. But, the local interaction effects also appeared to be a significant determinant, and the worker heterogeneity and the local interaction effects together account for about two-thirds of the wage disparities among employment areas. Mion and Naticchioni[120] provided similar evidence regarding the relationship between the spatial wage disparities and worker characteristics as well as market potential for the case of Italy using the similar large panel data.

---

16 See Section 5.1.
17 See also Mion [119] for a recent extension to Hanson[75]’s work.
18 In particular, their “market potential” is closer to Harris[76]’s classic definition, i.e., the spatial proximity to the total population.
4.2 Agglomeration shadow and spatial coordination of industrial agglomerations

Transport costs (broadly defined) matter in determining the spatial distribution of economic activities, and thus, interregional and international trade. While this fact is supported by several empirical results,\textsuperscript{19} it is not specific to the NEG. As far as the transport costs, or distance, is concerned, the specific theoretical contribution of the NEG is to provide a microfoundation for the agglomeration shadow (refer to Section 2.2). Distance matters in such away that agglomerations of each industry (roughly) take place at a certain industry-specific distance, where the area between agglomerations is in the agglomeration shadow of the industry, and thus not a profitable location. Moreover, while the spacing of agglomerations differs across industries, there is a mechanism in the NEG framework (e.g., Fujita, Krugman and Mori[60]) that the locations of agglomerations of different industries pull each other, and lock-in via demand/production externalities. So that more localized industries (with a smaller number of agglomerations) tend to agglomerate on top of the agglomerations of less localized industries (with a larger number of agglomerations), constituting a hierarchal structure in which the location with a more localized industry has all the less localized ones as well. If the location of industrial agglomeration is considered to represent a metro area, this result suggests the hierarchical industrial structure among metro areas where naturally a larger metro area has a larger variety of industries, which contains the set of industries located in smaller metro areas as a subset.

To date, very few attempts have been made to identify the spacing of agglomeration and the spatial coordination of agglomerations across industries. Here we review a few related works.

The attempt by Ioannides and Overman [85] was the earliest to test the empirical relevance of the NEG. They investigated, in particular, the role of the distance between agglomerations and the impact of market potential on the size and growth rate of cities. While they obtain only a few robust supportive results, it may be due to their inaccurate specifications for important variables such as distance and market potentials. For instance, the definition of their market potential is closer to that of the traditional Harris’s [76], and not specific to each industry. The NEG suggests that the market potential is specific to industry, and it explained the different sizes of agglomeration shadow, and hence, the different spacings of agglomerations of different industries. Also, the distance in their data is geodesic distance. Since the topography of the US is quite inhomogeneous (e.g., road network in the East and Mid-West is much denser than the West and South due basically to topographical reasons), it should be somehow controlled.

Duranton and Overman [48], by utilizing the inter-establishment distance

\textsuperscript{19}See Anderson and van Wincoop [8] for a comprehensive survey of empirical evidence for the significance of trade costs. In particular, it is reported that the tax equivalent of iceberg trade costs for industrialized countries is 170 percent, among which the transport costs account for 21 percent.
data, developed a statistical method to identify the spatial extent of industry-specific as well as inter-industry agglomerations.

More direct empirical evidences for the spatial patterns of industrial agglomerations *a la NEG* have been provided by Mori, Nishikimi and Smith [123][124][125]. Mori, Nishikimi and Smith [123] developed a geographically-decomposable test statistic of industrial localization based on the regional data. By computing the regional decomposition of the localization for the case of Japan, they have shown that the localization across metro areas explains on average more than 40% of the total degree of localization (measured at the county level). Moreover, they found evidence for the hierarchical structure in the industrial composition among metro areas (i.e., the set of industries found in a smaller metro area is roughly a subset of those found in a larger metro area). Mori, Nishikimi and Smith [124] provided alternative test for this hierarchical structure, and have shown that even after controlling for the industrial diversity of each metro area, and for the number of agglomerations of each industry, this hierarchical structure appears to be robust.

Mori, Nishikimi and Smith [125] was the first to provide a test of the presence of agglomeration shadow. To do this, they first developed a statistical method to identify each industrial agglomeration, and then tested if the distance between the neighboring agglomerations of the same industry is significantly larger than what would be realized if the same number (and areal size) of agglomerations are randomly generated. For the case of Japan, it has been shown that the agglomeration shadows are more clearly detected for less localized industries. Moreover, based on the distance from each agglomeration of a given industry to the nearest agglomeration of another industry, they constructed a test of spatial coordination of agglomerations between each pair of industries. Their preliminary results show that most market-oriented (or second-nature based) industries exhibits strong positive coordination (i.e., their agglomerations pull each other), while those based on the first nature such as the location of natural resources shows little coordination with other industries.

### 4.3 Home market effects

When production is subject to increasing returns, there is an incentive to concentrate this production in only a small number of locations. Hence the (original) version of the HME: the region with the relatively large number of consumers is the net exporter, and hosts a more than proportional share of firms in the increasing returns sector. In the presence of transport costs, these locations tend to be closer to large markets to save transport costs. Hence, the spatial version of the HME: other things being equal, each region tends to export goods and services for which this region has not only larger domestic market, but also a greater proximity to foreign markets. For the empirical evidences in terms of the non-spatial (original) version of the HME have been surveyed by several authors.\(^{20}\) Thus, here we only introduce a new (and the first) contribution re-

\(^{20}\)See, e.g., Anderson and van Wincoop[8].
garding this spatial version. Behrens, Lamorgese, Ottaviano and Tabuchi [22] developed a testing framework for the HME in the multi-country economy by extending the model by Krugman [102]. In their model, the advantage of a country as a production site depends on both the relative size of its domestic market and its relative proximity to foreign markets. They have shown that this extended model predicts a HME only after the actual production and trade data have been corrected for the impact of countries differential access to world markets, and in fact confirmed the strong evidence of HME in the world trade data.

4.4 Cumulative process and multiple equilibria

In the NEG framework in which there is no a priori advantageous location, the history and the cumulative process determine the resulting location of economic agglomeration. In such case, multiple (even continuum of) equilibria may prevail. It is not difficult to find the evidence for this path-dependence for an individual industrial localization. But, it is not at all obvious for the case of agglomeration of overall economic activities, or the formation of cities.

A recent work by Davis and Weinstein [41] looked at the effects of the US bombing campaign against Japanese cities, and found that the extent of damage had no effect on the population size of a city once the postwar recovery was fully achieved. That is, prewar population has correctly predicted postwar population, independent of the damage in between. By this result, they conclude that there is no evidence of cumulative effects of temporary shocks (bombing) which we should have expected if multiple equilibria prevail. Davis and Weinstein [42] reexamined their earlier result in a more strict context, and still found no strong evidence for multiple equilibria. But, this evidence is not inconsistent with NEG if the location space is not homogeneous (e.g., Fujita and Mori [62]). Namely, if some locations are given natural advantage such as natural ports with large hinterlands just like many locations of Japanese cities, then it is not surprising that there is strong tendency for economic agglomerations to take place at these locations.

On the other hand, Rhode [147] found evidence of a cumulative process that the economic boom in the Pacific Coast during the second world war boosted California and the West into a higher level equilibrium which persisted thereafter.

The research in this topic is very scarce. These inconclusive evidences indicate an urgent need for more empirical studies on this topic.

5 The way forward

We close the paper by listing up a few promising directions for further research based on the NEG models.
5.1 Unifying urban economics and the new economic geography

Urban economics and the new economic geography are largely treated today as two distinct fields, although they deal with essentially the same spatial phenomena. There are two major difference in the model architecture of these two fields.

One is the source of dispersion force. On the one hand urban economics models consider land rents for urban housing (and commuting) as a dispersion force. But, cities in these models are “floating islands,” where the intra-city and inter-city spaces are not integrated in the same location space. In fact, in the two region framework, later refinements of the core-periphery model (Helpman[78], Tabuchi[152], Murata and Thisse [132]) by incorporating the land rent for urban housing (and possibly commuting) as the dispersion force have shown that when transport costs of manufactures become sufficiently low, the industry disperses again to the periphery unlike the original Krugman’s model. However, given the discrete nature of space, it is difficult to distinguish whether such an industrial dispersion corresponds to the formation of many cities (as in Fujita and Mori[63], Fujita, Krugman and Mori[60], Krugman[98]), to the formation of an industrial belt (as in Mori[121]), or to merely a suburbanization within a metropolitan area (as explored in Fujita and Ogawa[65] and Ogawa and Fujita[133]). Notice, however, that simply incorporating many regions would not solve this problem if the interregional spatial structure is symmetric, i.e., the distance between any pair of regions is the same as in Tabuchi, Thisse and Zeng [154]. Since all the regions are geographically “next to each other,” just as in the two-region model, the meaning of “dispersion” into a larger number of regions is not clear. In order to investigate the spatial pattern of agglomeration, the “asymmetry” rather than symmetry of the location space is necessary where not all other regions are neighbors of each region.

On the other hand, the models in the early stage of the NEG framework (Krugman[96], Fujita and Krugman[58], and their direct extensions) considered the immobile resources (such as land) as the source of dispersion force, and by doing so focused on the spatial distribution of cities, while abstracting from the intra-city structure [i.e., a city consists of a (spaceless) point in the location space]. In order to investigate the impact of urban costs on the spatial pattern of agglomeration, we need to unify the NEG models and traditional urban models, and to study both the development of cities (having spatial extent) and industrial agglomeration in the same continuous space.

The other distinction of the two fields is that urban economic models assign big roles to developers and city governments, while the NEG has been concerned with self-organization in space while neglecting developers and governments. An obvious next step is to cross-fertilize the two fields.

Indeed, some efforts have been undertaken along this line in recent years. For example, the introduction of Tiebout-type inter-jurisdictional competition (in taxation and the provision of local public goods) into core-periphery models has
been achieved recently by several authors such as Baldwin and Krugman \cite{14}, Kind, Midelfart-Knarvik and Schjelderup \cite{92}, and Anderson and Forslid \cite{7}. The next step would be to graft various urban features (such as land and housing markets, commuting, transportation networks and other urban infrastructure) onto geographical models with local governments. In modeling the competition among cities, however, we should note that most city-governments / developers have, in practice, very limited powers and foresights in choosing their policies. Hence, instead of full-fledged Nash games, it may be more appropriate to use a certain kind of evolutionary games combined with an appropriate political process such as voting.

Eventually, we must squarely face the concept of cities as seedbed for the generation, diffusion and accumulation of knowledge, which is central for the phenomena of innovation and economic growth.\textsuperscript{22}

5.2 Morphology and spatial distribution of agglomeration

After all, to date only very limited low-dimension problems—mostly in a setup with two regions and a few sectors—have been solved by pencil-and-paper analyses, and much more interesting issues related to, for instance, the spatial distribution of agglomerations (e.g., the number, size, location and spatial coordination of industrial agglomerations) and the form of agglomeration/dispersal (e.g., if dispersion means suburbanization, megalopolis formation, or emergence of many “discrete” cities, etc.) are left untouched after the initial (informal) explorations by Krugman\cite{98,101}, Krugman and Venables\cite{106}, Fujita and Mori\cite{63,62}, Fujita, Krugman and Mori\cite{60}, and Puga and Venables\cite{143}. In particular, as discussed in Section 5.1, under two-region setups, it is not often possible to distinguish between qualitatively different forms of agglomeration and dispersion.

This situation would not probably improve so soon. For the past decade, the use of numerical simulations became rather minor in the NEG research. However, given the ever improving power of computer hardware and more importantly, the recent popularization of sophisticated object-oriented programming environment,\textsuperscript{23} it may be time to revisit the possibility of computable geographical equilibrium models.

While it will continue to be important to pursue building analytically solvable models regarding the basic mechanism of agglomeration and dispersal, it will become even more important to build numerically computable models. After all, there is great need to finally go beyond the basic two-region-two-industry models and go to asymmetric many-region-many-industry models of trade and geography in order to draw practically useful policy implications. Most models with emphasis on the analytical solvability are solvable only in a very limited low-dimensional setup, but they are often not computable numerically (at least

\textsuperscript{21}For a review, see Baldwin et al. \cite[Ch.15,16]{16}.

\textsuperscript{22}See Duranton and Puga\cite{45} for a recent development.

\textsuperscript{23}The standardized (and accessible) version of the most popular object-oriented programming language, C++ (see, e.g., Stroustrup\cite{151}), was made available after the first generation of the NEG works had already been produced.
in a practical amount of time) once more general spatial and industrial structure are incorporated. A most desirable model would be the one that have solvability at the low dimensional setup and computability even at the fairly high dimensional setup.24

5.3 Transportation technology and agglomeration

Industrial agglomerations often appear in association with major transport nodes.25 Obvious examples are those in cities which are usually seen near key junctions of highway networks or large railroad stations. At a more aggregated level, the unprecedented growth in Asian industries in the 1980s took place around the three largest ports in the world: Hong Kong, Singapore, and Kaohsiung. Major transport nodes have continued attracting industrial locations even though there has been a substantial fall in transport costs.26

The coincidence of the industrial agglomeration and transport nodes results from the process of reciprocal reinforcements between them. Of the two reinforcement forces, firms’ motivation to save transport costs attracts these firms to locate around transport nodes. Indeed the total transport costs paid by major manufacturing firms in Japan amount to 8.69% of their total sales value (Japan Logistic Systems Association[87]). Anderson and Wincoop[8] estimate that the tax equivalent of trade costs for industrialized countries amounts to as large as 170%, which roughly decomposed into transport costs (21%), border related trade barriers (44%), and local retail and distribution costs (55%).

In addition to these pecuniary costs, firms bear significant time costs for transportation. In particular, they often need business contacts with their customers and material suppliers in other regions. Even within a firm, local managers must regularly meet to discuss business decision. All these things, of course, require frequent business trips across regions which incur a lot of time and money. For another example, assembly firms of electronic products in Asia are constantly subject to uncertain changes in market demand and production technologies. They are thus forced to frequently alter the amount and variety of components to be assembled. If the transportation of components takes time, they need to order them much earlier without knowing the exact type

24For instance, Ottaviano[135, Ch.3] and Mori and Turrini[126] propose such “modified” NEG models. Although their original models are formulated in two-region setups, the complexity of their models does not explode even after introducing many regions and many monopolistically competitive industries. For instance, the number of real unknowns for a temporary equilibrium in their model is independent of the number of industries, but only depends on the number of regions that have positive agglomeration of some monopolistically competitive industries. Such computability is not present in the original NEG models such as Krugman[96][98] and Fujita and Krugman[58].

25As Hakimi[73] and its extention by Louvex, Thisse and Beguin[108] have shown that the nodes (and transshipment points) of transport network always contain an optimal location for cost-minimizing firms.

26For instance, Glaeser and Kohlhase[?] reported that the share of transport industries in the US GDP has decreased more than two fold during the 20th century. Similarly, Combes and Lafourcade[?] computed the decline of 40% of the cost of road freight between 1978 and 1998.
and amount of necessary components. To avoid this sort of risk, they prefer to operate at locations with good transport access, such as large international ports.

The other reinforcement force is that the efficiency of transport nodes is improved by the increase in transport demand stemming from the growth of industrial agglomeration. The basic mechanism originates from scale economies in transportation which have been realized by the development of large-sized and high-speed carriers, such as container ships, bullet trains, and jumbo jets. The scale economies provide an incentive for collective transportation and hence stimulate the development of trunk routes and the hub-spoke structure of transportation. The process of the trunk route formation exhibits the following circular causation. Suppose there are frequent transport services on a given link, such that these are available on demand. As a result, a large number of shippers are attracted to use the link, which in turn supports even more frequent transport services on the link. This positive feedback mechanism eventually leads to the endogenous formation of trunk links and transport hubs. When scale economies in transportation rule the transport advantage of each location, a major transport node can spontaneously emerge at any place having large transport demand like the location of industrial agglomeration. We call the above mechanism of circular causation economies of transport density. Several studies have shown evidence that economies of density are significant in air, railroad and maritime transportations.

There are mainly two groups in the existing literature on the causal relationship between industrial location patterns and the transport network structure. The works in one group depict the design of a transport network as a problem of a planner in a transport sector when economies of density exist. However, origin-destination flows between each pair of locations are assumed to be given in their models. As a result, they do not explain how the structure of the transport network affects the industrial location pattern. On the other hand, the works in the other group focus on deriving industrial location patterns under a given structure of the transport network. A few attempts using the NEG framework belong to this group (Fujita and Mori, Krugman and Mun). However, they do not explain how the spatial distribution of economic activities affects the structure of transport network.

Behrens has pointed out the iceberg transport cost in the NEG itself implicitly assumes some form of increasing returns in transportation which tends to generate clustering of firms and consumers at a point on the location space in the NEG models. The first attempt to model the transport sector explicitly in the NEG framework was Takahashi who developed microfounded formulation of economies of transport density endogenizing the transport sector in the

---

27 See Brueckner and Spiller, Brueckner, Dyer and Spiller, and Caves, Christensen and Tretheway for the case of air transportation, and Braeutigam, Daughety and Turnquist for the case of rail transportation, and Mori and Nishikimi for the case of maritime transportation.

28 See, e.g., Campbell and Hendricks, Piccone and Tan.

29 See also Konishi and Mun for models outside the NEG framework.
context of the two-region setup. Due to the two-ness of the location space, however, there is no network or hub formation in this model, thus interdependence between the agglomeration patterns and the structure of transport network is yet to be explained.\textsuperscript{30} Further development in this direction is obviously among the most immediate agenda.

Finally, we may note that in the actual application of NEG models, it is important to distinguish the two different types of impediment to trade in space, i.e., transport costs for goods and communication costs for doing business over space. The recent works by Fujita and Thisse \cite{70} and Fujita and Gokan \cite{57} suggest that transport costs for goods and communication costs among business units exert different effects on the spatial organization of economic activities.\textsuperscript{31}

In short, comprehensive studies on the nature and modeling of transportation activities should be, of course, a central concern for the further development of the NEG.

5.4 Linkages through the creation and transfer of knowledge in space

In most models of the NEG so far, agglomeration forces arise solely from pecuniary externalities through linkage effects among consumers and industries, neglecting all other possible sources of agglomeration economies such as knowledge externalities and information spillovers. This has led to the opinion that the theories of the NEG have been too narrowly focused, ignoring as much of the reality as old trade theory.

We fully understand the concern. But, first, let us defend our position. It is true that the theoretical framework of the NEG has been very narrowly focused. But, it was a deliberate choice. That is, such a narrow focus of the NEG was designed in order to establish a firm micro-foundation of geographical economics based on modern tools of economic theory. It does not necessarily mean that the NEG is limited to such a narrow range of models and issues. On the contrary, its framework is widely open to further development. Indeed, recently many of such possibilities are being explored vigorously by many young scholars, as some of them have been reviewed in this paper.

That much said, however, we admit that there still remains a big room for further development of the NEG. In particular, there remains one type of agglomeration forces of which micro-foundations have seen little development so far: that is, the linkages among people through the creation and transfer of knowledge, or in short, the K-linkages. (Hereafter, "knowledge" is defined broadly to include ideas and information.)

\textsuperscript{30} Even outside the NEG, Mori and Nishikimi\cite{122} remain to be the only model which explains the interdependence between industrial location and transport network formation (i.e. hub formation) incorporating the economies of transport density.

\textsuperscript{31} Duranton and Storper\cite{49} provide a model explaining the situation in which transport cost for distance falls while it triggers more frequent interaction across space, and ends up with an increase in total transport costs.
Traditionally, K-linkage effects have been called either "knowledge spillovers" or "knowledge externalities". However, the term, "spillovers", tends to have a connotation of passive effects. And, the term, "externalities", tends to imply too many different things at once. So, in the remaining discussion, instead of knowledge spillovers or externalities, we use the term, K-linkages, in order to emphasize that they represent the agglomeration forces resulting from the activities related to both the "creation of knowledge" and the "transfer of knowledge" or "learning" (either in an active way or a passive way). In contrast to the K-linkages, the traditional linkages through the production and transactions of (traditional) goods and services may be called the "E-linkages" (where "E" representing the economic activities in the traditional economics).

Using such a terminology, we may imagine that the agglomeration forces in the real world arise from the dual effects of E-linkages and K-linkages. In this context, we conjecture that the role of K-linkages has been becoming increasingly more dominant recently. Yet, developing the micro-foundations of K-linkages seem to be the most challenging task, largely left for young scholars in the future.

We are in haste to add that there has been a great amount of conceptual studies on knowledge externalities / spillovers in a spatial context, starting with Marshall [112], and including more recent pioneering work such as Jacobs[86], Anderson[6] and Lucas[109] in an urban context, and Porter[140] in the context of industrial clusters. Yet, it would be fair to say that there is a lot of room left for advancing the micro-foundations of K-linkages in space. Particularly, in developing the micro-foundations of K-linkages, "creation of knowledge" must be clearly distinguished from "transfer of knowledge" or "learning". Furthermore, for the creation of new ideas, cooperation among heterogeneous people are essentially important. Yet, through communication and migration, the degree of the heterogeneity of people in a region changes over time. Thus, the nature of K-linkages is essentially dynamic, and hence their full-fledged treatment requires a dynamic framework.

Recently several pioneering works have appeared on the dynamic models of K-linkages, although they are mostly aspatial. Among others, Jovanovic and Rob[90][89], Jovanovic and Nyarko[88], Auerswald et. al.[9], and Keely[91] present micro-models of K-linkages. These studies are mostly concerned with the dynamics of vertical differentiation of knowledge. In contrast, Berliant Reed and Wang[26] and Berliant and Fujita[25] are concerned with the dynamics of heterogeneous knowledge differentiation through the cooperative processes of knowledge creation and transfer.

Building upon such pioneering works, it is hoped that micro-foundations of K-linkages in space will be developed in the near future. Then, we may be able to develop a comprehensive theory of the NEG, which integrates fully the dual effects of E-linkages and K-linkages in space. Using such a generalized framework, for example, we may be able to explore the economic implications of cyclical migration of skilled workers (suggested in Section 3.5) on the interregional transfer of knowledge and skills as well as on the knowledge-heterogeneity within each region and among regions.
References


25


28


