Spatial implications of international trade under the new economic geography approach

Ramírez Grajeda, Mauricio and de León Arias, Adrián
Universidad de Guadalajara, Universidad de Guadalajara

01. October 2009

SPATIAL IMPLICATIONS OF INTERNATIONAL TRADE UNDER THE NEW ECONOMIC GEOGRAPHY APPROACH

Adrián de León Arias[a] & Mauricio Ramírez Grajeda[b]

Abstract

In 2008, Paul Krugman from Princeton University was awarded the Nobel Prize in Economic Sciences by the Central Bank of Sweden, for his "analysis of trade patterns and location of economic activity". In this paper we survey the literature, known as the New Economic Geography (NEG), launched by Krugman (1991). In particular, we focus on four topics: *(i)* NEG roots, *(ii)* NEG rationale; *(ii)* the spatial impact of international trade on global economic imbalances; and *(iv)* the impact of international trade on urban structure.

***Keywords**:* New Economic Geography, Trade Openness, Agglomeration and Urban Economics.
JEL Classification: F12, F15 and R12.

Introduction

According to Venables (1998), a key question for the future development of the world economy is, how global integration impacts on the location of economic activity? In particular, what is the effect, at international and regional level, of international trade openness on the spatial pattern of production, welfare and trade? This question, for example, was in the center of the political debate over the North American Free Trade Agreement (NAFTA). Hanson (1998) points out that most of the U.S. congressional representatives from districts near Mexico strongly supported NAFTA, while those ones from districts close to Canada offered resistance. Such an attitude toward NAFTA reflects the perception that firms would move away from northern states to the south to reach new markets. Another example, given by Venables (1995), is the concern about the spatial implications of the European Union (EU) enlargement by the end of 2004.

By using the Ricardian comparative advantage theory, especially the widely employed Heckscher-Ohlin-Samuelson version, as a standard tool, we could find that the determinants of spatial patterns of production would be based on differences in factor endowments, technologies,

[a] *Corresponding author*. Universidad de Guadalajara, Centro Universitario de Ciencias Económico Administrativas, División de Gestión Empresarial. Periférico Norte 799, Módulo E, primer nivel, C.P. 45100, Zapopan, Jalisco, México. *E-mail:*leonarias@yahoo.com..mx. Tel. +52 (33) 3770 3300 ext. 5062.
[b] Universidad de Guadalajara, Centro Universitario de Ciencias Económico Administrativas, Departamento de Métodos Cuantitativos. Periférico Norte 799, Módulo M, segundo nivel, C.P. 45100, Zapopan, Jalisco, México. *E-mail:* ramirez-grajeda.1@osu.edu. Tel. +52 (33) 3770 3300 ext. 5223.
Conacyt partially supported this investigation. We thank Paulina Brambila Trejo and Antonio Romero for their assistance.

preferences or trade policies. Therefore countries or regions would specialize according to their comparative advantage. However, Venables (1998) states that an issue that cannot be adequately addressed under this theoretical framework is the location of economic activity across countries or regions, where endowments are broadly similar (eg. the EU) or within which factors are mobile (eg. the US). The conventional trade theory would predict that economic activity should be uniform but this is not the case. Even more, there is not convergence.

Although this topic is inherently very important, Krugman (1979) and Fujita et al. (1999) consider that until the early 1990s geographic considerations have been neglected by mainstream economics. For example, Krugman and Livas (1996) claim that in geographic production concentration as the growth of large cities has been obliquely addressed the development economics literature. Fortunately, there is a long tradition of analysis in spatial economics and, as Krugman (1998b) recognizes, spatial economics has received considerable attention in recent years. Both factors have motivated a theoretical approach known as NEG, which provides an interesting framework to answer our initial questions. This concurs with Ottaviano and Thisse (2004) for whom:

> "… many of the NEG ideas have been around for a long time in the works of economic geographers and location theorist. However, NEG has the fundamental merit of having framed those ideas within a general equilibrium model encompassing most of these ideas. This has drawn economic geography and location theory from the periphery to the center of mainstream economic theory. More importantly, it has made already existing ideas more amenable to empirical scrutiny and policy analysis."

Krugman (1991), Fujita (1993) and Venables(1996) are regarded as having given birth to the NEG paradigm, which uses full-fledged general equilibrium models with monopolistic competition à la Dixit and Stiglitz (1977). The NEG literature could be divided according to two mechanisms of agglomeration. One is allowing labor mobility, which is a distinctive feature at regional level. The other is incorporating backward and forward linkages but impedes labor mobility, which is a distinctive feature at international level.

At regional level, Krugman (1991) and Fujita (1993) show that the combination of increasing returns to scale, trade costs and the mobility of industrial labor force creates a feedback process of industry agglomeration. Advantages (or centripetal forces) for firms of being close to large product and labor markets arise from reductions in trade costs and nominal wages. Moreover, agglomeration attracts workers because it induces a wider product variety and higher real wages As more firms locate in one production site, this one turns out to be more profitable. Yet, there are disadvantages (or centrifugal forces) of agglomeration. Firms face more competition in the product and labor

markets and workers face higher congestion costs. Therefore, there is room for agglomeration as long as advantages generated by it outweigh its disadvantages. Metropolises like Tokyo, Sao Paulo or Bombay are examples of cities where agglomeration equilibrium has not been reached yet.

At international level, Krugman and Venables (1995), Venables (1995) and Puga (1999) show that without labor immobility backward and forward linkages happen to be a mechanism of agglomeration: firms produce and purchase inputs. For firms clustering means lower input costs and a larger product market. International trade costs determine the importance of both linkages in location decisions.

Here, it is noteworthy that two elements are essential in firms´ location decisions: trade costs and increasing returns to scale. The former drive firms to supply near large markets; the latter drive firms prefer to serve from a single location. Almost all the initial ideas on location theory assume economies of scale, which enforces geographic concentration of economic activities. For example, within the German tradition of Weber (1909), Christaller (1933) and Lösch (1954); an exception is Von Thünen (1826). Yet such works are developed within a partial equilibrium framework. In a general equilibrium context, by assuming both constant returns to scale and positive trade costs it is hard to realize why the economy does not fall into a Robinson Crusoe type, where each household produces her own consumption. This result is known as the "folk theorem" of spatial economics. Thus it is fair to say that spatial issues make more sense with increasing returns to scale and positive trade costs.

For Brakman et al. (2001), psychological, sociological, cultural and historical forces are behind spatial clustering. Although these are valid perspectives, in this paper we review the literature related to the impact of trade openness on geography under the NEG approach. But first survey the NEG intellectual background. In particular, we devote our attention to the main ideas out of economists, geographers and regional scientists that have contributed to build the NEG ideas. The novelty of this paper is that it covers theoretical and empirical papers regarding international trade.

It is worth mentioning some surveys have also focused on the NEG background.[1] Brakman y Garretsen (2009) argue that Krugman (1991) is closely linked to Krugman´s (1979, 1980) trade theories. Fujita and Thisse (2009) relate both the Urban Economics literature and Location theory to the NEG framework. Ottaviano and Thisse (2004) pay attention on the main contributions of location theory by geographers and regional scientists to NEG. They divide their analysis in two parts: The location of firms as a result of an individual decision, and the location of the industry as a result of firms´ interactions. Head and Mayer (2003) examine empirical strategies to test NEG

[1] Some of these surveys do not focus on NEG background entirely.

features and predictions. Ottaviano and Puga (1998) focus on comparative advantage and market access considerations to explain the spatial distribution of economic activity. Venables (1998) reviews the old tradition of development economics and regional economics to link spatial agglomeration and cumulative causation. Krugman (1998a, 1998b) states that old ideas on spatial economics were neglected by mainstream economics due to technical obstacles. The main one was the impossibility to fit a model with increasing returns to scale. Quigley (1998) links urban diversity and economic growth and presents a chronological description of this issue. Fujita and Thisse (1996) present the main contributions of location theory and standard economic theory to NEG.

The reminder of the paper is divided up as follows. Section 1 presents NEG intellectual roots. Section 2 provides the economic rationale of the NEG paradigm. In section 3, we survey the literature related to Puga´s (1999) remarkable theoretical outcome: industrial concentration has a bell-shaped[2] relationship with international trade costs. Our aim is not only to describe the most relevant contributions of these specific research lines within trade theory, but to present the main technical aspects of such literature. In particular, we pay attention to their assumptions, mathematical tricks, unrealistic results and empirical test possibilities. In section 4**,** we cope with the effects of trade openness on cities´ size. The NEG literature related to this issue is relatively scarce. It is divided into two parts: One refers to theoretical works; the other refers to empirical research. Finally, some implications of our survey, in terms of main NEG shortcomings and the way forward, are presented.

1. Intellectual Underpinnings

Constant returns to scale imply that activities are divisible, thus each activity can be carried out at any scale without sacrificing efficiency. Thus, autarky is a competitive equilibrium involving positive trade costs. However, Starrett (1978) proves that if indivisibilities are assumed instead, then there is no a competitive equilibrium. So understanding spatial patterns requires deviating from Starret´s (1978) setting. As Fujita *et al.*(1999) proceed, we identify the antecedents of NEG into three alternative theories on industrial location: Marshall-Scitovsky externalities, urban economics and regional science.

[2] It is also known as the U-shaped or inverted U-shaped curve.

Marshall-Scitovsky Externalities

Economic agglomeration could arise as a consequence of the presence of externalities. In the literature there are two dominant points of view with respect to externalities. On the one hand, Marshall (1920) explains different ways in which industry's output as an argument of firms´ production function foster agglomeration, such as, informational spillovers that expand firms´ production set when they cluster together; access to thick consumers and inputs markets, as well as the formation of high skilled labor force based on the accumulation of human capital and face-to-face communications. This ensures that both unemployment and labor shortage is unlikely. On the other hand, Scitovsky´s (1954) externalities can be divided into two categories: technological externalities and pecuniary externalities. Technological externalities refer to the direct impact of production and consumption activities on production and consumption sets. The market structure associated with this type of externality is perfect competition. Ottaviano and Thiesse (1997) consider that an example of this type of externality arises in certain location if the arrival of new firms increases the efficiency of local firms because they enhance the productivity of labor through social learning process. Pecuniary externalities are the benefits of economic interactions that are transmitted through market prices. The market structure associated to pecuniary externalities is imperfect competition. Ottaviano and Thiesse (1997) consider that an example of pecuniary externality could be the reduction of output prices due to additional supply generated by the inflow of new firms in certain location.

It is worth noting two points. First, Marshallian externalities turn out to be a combination of technological externalities and pecuniary externalities. Second, as a result of the first point explained above the market structure associated with each of the Marshallian externalities is not straightforward[3]. Externalities seem to be unrelated to agglomeration, but they are essential ingredients of the NEG rationale.

Urban Economics

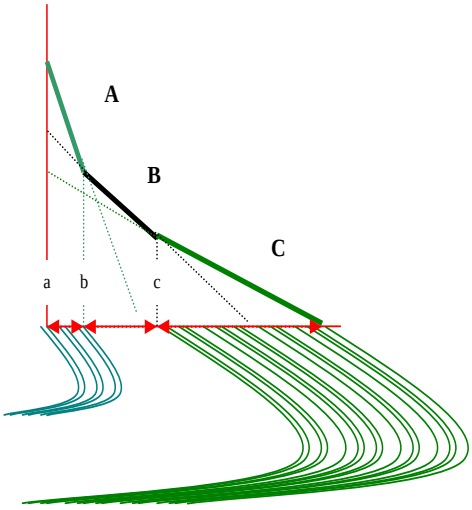
Fujita et. al. (1999) point out that urban economics is a branch of the economics which has been forced to take spatial concerns into consideration. Von Thünen (1826) is a pioneer model in urban economics, and it remains as a benchmark to this day for its clear exposition of land use surrounding a city. It is worth mentioning that this model does not rely on scale economics. His setting assumes the existence of a plain which is homogenous in every attribute. In this plain there

[3] Helpman ad Krugman (1985) explain the problems to associate a market structure with each of the Marshallian externalities.

is a single urban center. Outside the urban center, agricultural producers sell their crops in the city. There are positive trade costs associated with transporting agricultural produces to the city, which differ for the various crops. The prices for these crops might also differ. The model analyzes how the farmer determines her location across the plain. Each farmer wants to be as close to the city as possible to minimize trade costs. The motivation to be close to the town pushes land rents up near the city. Each farmer thus faces a tradeoff between land rents and trade costs.

Von Thünen (1826) shows that competition for location ensures that the equilibrium allocation of land among the agricultural producers is efficient. For every type of crop there is a bid-rent curve which indicates, as a function of the distance to the city, how much farmers are willing to pay for the land. Bid–rent curves differ by crops due to different given prices for those crops and their respective trade costs. It turns out that the farmers of a particular kind of crop are able to outbid their competitors for any given distance to the city. In figure 1, we can observe that as producers move away from the city center, producers of A outbid the other two groups of farmers. Between points b and c, producers of B are willing to pay the highest rents; Beyond c, producers of C pay the highest rents.

Figure 1. Bid-rent curves



Von Thünen (1826) has an important reappearance in Alonso (1964) who reinterprets it by substituting commuters for farmers and a central business district for an isolated city. This model again yields concentric rings of land use, and it is a seminal paper for an extensive theoretical and empirical literature on urban sprawling.[4] Nevertheless, such a framework has an important shortcoming: the existence of a town or a business district is simply assumed. It does not answer

[4] We recommend Brueckner (2000) as an excellent introduction to urban sprawling.

how land use is determined when the location of the town is itself endogenous. The Von Thünen (1826) model is complemented with the concept of externalities in subsequent literature.

Regional Science

Weber (1909) is also a partial equilibrium model, which frames the problem of location in terms of an individual producer who takes the locations of other producers and all prices (including her own) as given. Subsequent work has enlarged on this, notably by letting prices be endogenous, and by considering strategic considerations of location decisions from different firms. Nonetheless, the geographical distribution of demand, and the location of inputs sources outside the industry in question are given.

Christaller (1933) and Lösch (1954)[5] explain the location of cities and differentiate cities by the various functions they perform. Both works assume that agents are evenly distributed across a featureless plain; the supply of goods and services consumed by the agents involves increasing returns to scale with positive trade costs. Central places serving the surrounding agents arise as a result of the trade-off between trade costs and scale economies. Christaller (1933) points out that this will create a hierarchically organized large number of market towns. A large city will produce all types or varieties of goods; small cities that cluster around the large one will produce a limited amount of varieties of goods; and the variety of goods produced by villages, which are around small cities will be even less than small cities. Lösch (1954) concludes that the form of this hierarchical system will be hexagonal. This story can be understood at many levels. For example, small districts could be scattered around larger districts, all eventually centering on the downtown. A shortcoming with this approach is that their exercise (rather than a causal model) describes planning solutions rather than market outcomes. The economic rationale behind the actions of firms and consumers are not addressed. The German school was aware that their story lacked of optimizing agents and general equilibrium considerations.

According to Krugman (1998a), the idea that agglomeration involves a circular process is not new. Harris (1954) and Pred (1966) develop a model in which firms choose locations with good access to markets and suppliers. This decision improves the access to market and suppliers. Krugman (1998b) considers that Harris (1954) and Pred (1966) provide a coherent and intuitively compelling story about urban agglomeration.

In Harris (1954) firms produce in sites with "market potential" defined as

[5] Their work constitutes what is known as central-place theory.

$$MP_i = \sum_{j=1}^{R} \left(\frac{M_{ij}}{D_{ij}} \right),$$

where M_j is the demand of location j for goods of location i. D_{ij} is the distance between locations i and j. Using the value of retail sales per U.S. county, his results show that highly industrialized regions were also locations with high "market potential". This supports the notion of clustering of economic activity is driven not only by the supply side but by the demand side as well. Under this result Harris (1954) suggests that production is self-reinforcing. Firms tend to produce in regions with high "market potential"; and "market potential" of regions tend to be higher in locations where firms decide to produce.

Pred (1966) is interested in the dynamics of regional growth by working with a simple "base-multiplier" model of regional income. The study starts with a projection of the export earnings of a region (its sales to other regions inside and outside the country), then uses an estimate of the share of income spent within the region to compute a multiplier on that base. Thus, if export income is $10 billion and 60 percent of such income is spent locally, then regional income will be 10/(1-.6)=$25 billion. Pred (1966) argues, however, that both the size of the export base and the share of income spent locally are increasing functions of the size of the economy. A sufficiently large scale economy could take off in a self-reinforcing dynamics of growth. For example, the large market might make profitable to produce locally goods that had been previously imported from other regions. This would increase the multiplier of the region's export base, leading to a further expansion of income, which would lead to still more local production.

2. The Rationale behind the New Economic Geography

The economic activities distributed across space can be explained, according to Overman et al. (2003), by using two spatial concepts: first-nature and second-nature geography. The former is the physical geography of coasts, mountains, and endowments of natural resources. The latter emerges as the outcome of agent's actions to overcome the constraints imposed by first-nature geography. Factor endowment-based trade theory considers the elements of the first-nature geography. Second-nature geography focuses on the implications of space and distance on agent´s behavior. NEG takes this second point of view after controlling for the first-nature. In this vein, the intuition behind the concentration economic activities is conceived as the outcome of two types of dynamic forces: Centripetal forces and centrifugal forces. NEG then combines and simplifies the ideas of Marshall (1920) and Scitovsky (1954) to formalize this intuitive explanation.

We can describe this formalization as follows. First, by defining specifically the centripetal forces, which are the Marshallian externalities already explained. Yet, as Scitovsky (1954) points out, each one is formed by two components: pure externalities and pecuniary externalities. Then, in real terms we may say that there are six kinds of centripetal forces. NEG picks up a particular centripetal force: the pecuniary component of the market size Marshallian externality. For Henderson (2001), there are two sources that generate this externality: backward and forward linkages. The former arise when a location with high demand attracts firms to move there. The latter arise because a large local markets support the production of intermediate goods at low cost.

Second, we define the centrifugal forces according to Krugman (1998a): Immobility of factors as land, natural resources and, at international context, workers. Such forces drive against concentration of production. From the demand side, dispersed factors are positively correlated to consumers markets. Then producers have an incentive to move close to consumers. From the supply side, production must go where the factors are. Land rents drive up due to concentration of economic activity. Higher rents are a disincentive for agglomeration. And finally, concentration generates pure negative externalities such as congestion. NEG selects either factor immobility or congestions costs as a dispersion force.

For modeling strategy reasons, more than for empirical considerations, NEG has chosen those particular forces. Both forces create what Arthur (1989) calls "positive feedback" dynamics: production will tend to concentrate where there is a large market, but the market will be large where production is concentrated. This story, where agglomeration of economic activity is driven by two opposite forces is not new. De la Blanche (1921) explains the same idea; and, as we pointed out, Harris (1954) and Pred (1966) use this story as their central theme. Behind it was the assumption of increasing returns to scale at the firm´s level. Other papers also assume it as Weber (1909), that establishes that producer's location decision is the result of minimizing the combining costs of producing and shipping given that there is a single production site. Christaller´s (1933) and Lösh´s (1954) assumption are that some locations cannot support certain activities. In sum, agglomeration requires increasing returns of scale at the level of the firm. However, a space consideration also implies agglomeration costs. Otherwise, all the production will be set up in one location.

Table 1. Geographical Concentration Driving Forces

Centripetal forces	Centrifugal forces
Market size effects	*Immobile factors*
Thick labor markets	Land rents
Informational spillovers	*Pure external diseconomies*

Source: Krugman (1998a)

Krugman (1998a) asserts that this story was widely known in economics until 1990. Unfortunately, mainstream economics had paid little attention to most stories of location issues despite the fact of its simplicity and intuitive logic. The reason is that under economies of scale perfect competition is not feasible. In the 1950s and 1960s there were non-tractable models of imperfect competition. NEG consists of full general-equilibrium models, in which budget constraints on both money and resources are carefully specified. And the geographical distributions of population, demand and supply are all endogenous.

Spatial issues can be analyzed in two areas if we consider the centripetal forces that drive the formation of economic clusters of firms and households. First, informational spillovers under perfect competition by solely taking its pure externality component. Second, market size effects by solely taking its pecuniary component under monopolistic competition. We survey the second point. A third point of view arises when we consider spatial competition under strategic interaction. Hotelling (1929) is the seminal work to this third point.

According to Fujita and Thisse (1996), models related to pure externalities consider spatial equilibria under the influence of nonmarket interactions, which typically involve communication of knowledge, ideas and tacit information between agents (firms and/or household). For Ottaviano and Thisse (1997), these pieces of information constitute impure public goods that generate spillover effects from one agent to another. Informational spillovers models have been developed in urban economics with the aim to explain agglomeration of specific economic activities within a city or industrial district.

The models that consider market size effects like NEG are an adequate framework to explain interregional agglomerations such as the industrial distribution pattern in Europe. However, they can also be used to explain large metropolis as Krugman and Livas (1996).

Space finally made it into the standard economics because imperfect competition turned to be tractable. There are four revolutionary waves or phases that raised from imperfect competition models. The New Industrial Organization began with Dixit and Stiglitz (1977), which formalizes the concept of monopolistic competition by Chamberlain (1933). Both works develop tools that

triggered what is known as the New Trade Theory (NT) in the 1980s and the New Growth Theory and New International Economics in 1990s. Krugman (1991) is the seminal paper of NEG. In international theory, this framework has allowed international economists to explain intra-industry trade as Krugman (1980, 1981) do in a framework that is tractable and flexible to model imperfect competitive markets. A contribution to turn Dixit and Stigliz´s (1977) framework into a spatial model is the concept of iceberg type trade cost.

Chamberlain (1933) introduced the concept of monopolistic competition which is based in four assumptions. First, each firm produces at most one type of product. Second, each firm faces a downward sloping demand curve. Third, profits are zero. And fourth, a price change by one firm has minimum effects on the demand of any other firm's product. Under this framework there are non strategic considerations. Each potential firm faces a residual demand *D(pi; pj)* and a U shaped average cost curve. Equilibrium with free entry implies that the residual demand for each firm is tangent to its corresponding average cost curve. The quantity produced is less than the quantity that minimizes average cost or equivalently fixed costs are spread over few units. As Tirole (1988) this is more clearly when the average cost is defined as (fix cost+margina cost*q)/q, where the optimal quantity is infinite.

There is only one face of perfect competition. However, imperfect competition can be modeled in many ways depending on the assumptions and the issues to address as strategic behavior, preferences of consumers or type of good. Dixit and Stigltz (1977) has been a workhorse in many areas of economics. Nevertheless, as Chamberlain (1933), it is a very restrictive model in several assumptions: symmetry among varieties, the resulting absence of both monopoly rents in equilibrium, no strategic behavior of firms, homogenous technology and representative agent.

Quigley (1998) analyzes Dixit and Stiglitz´s (1977) model in its spatial version. On the consumption side, there is a representative household whose utility depends on traded goods, space and a variety of goods. The utility function exhibits constant elasticity of substitution. The market for traded goods and space are competitive, while differentiated local goods are sold in a monopolistically competitive environment. Variety and local goods positively affect household's utility. On the production side, variety of inputs has an equivalent importance as consumption. Production is a function of labor, specialized inputs and space. Labor and space can be taken as competitive, while inputs are traded in a monopolistically competitive market. Variety in inputs positively affects output (two inputs of different types produce more than two inputs of the same type). The amount of inputs and labor positively affects output. The main conclusion of this type of model is that variety of goods and inputs yield a dynamic and endogenous externality (centripetal

force). The size of the labor force in a particular location determines the variety of goods and inputs. The larger is a particular city, the larger the variety of goods and inputs is. The well-being of this city increases its size by attracting more labor force. This process stops as long as the centrifugal forces are weaker than the centripetal.

According to Krugman (1998a) any attempt to develop a general-equilibrium model of location would be substantially complicated by adding a transportation sector. To simplify the analysis, iceberg type transportation cost was first introduced by Samuelson (1954). It is the constant fraction of any shipped good that depreciates in transit between two places.[6] Under this type of costs the constant elasticity of demand is preserved.

NEG models rely heavily on ad hoc, although realistic assumptions. [7] Head and Mayer (2003) catalog five key ingredients of this paradigm:

1. *Scale economies at firm level.* Firms have fixed requirement for limited productive resources.

2. *Imperfect competition.* Provided the first ingredient, marginal cost is always lower than average cost, then perfect competition is an implausible market structure. NEG models are based on Dixit and Stiglitz (1977). In particular, consumer's love of variety, which is captured in by a CES utility function that is symmetric in a bundle of differentiated products; and the fact that firms have no influence on overall market conditions.

3. *Trade costs.* The outputs and inputs used by firms are tradeable over distances but only by incurring Samuelson type costs: A fraction of the good depreciates on the way. This assumption gets rid of having another industry and a variable demand elasticity.

4. *Endogenous firms locations.* Firms enter and exit freely in response to profitability at each location.

5. *Endogenous location of demand.* Both consumers and firms demand output. Such a demand work through two mechanisms that allow a cumulative causation process. Surveys as Ottaviano and Thiesse (2003) or Head and Mayer (2003) divide NEG models into two main directions according to the location of demand. One, is at regional level by assuming labor mobility across regions. The other, at international level by assuming labor immobility, and that industrial production requires the output of their sector as intermediate inputs.

[6] The concept of distance is not considered in NEG models. Then, transportation costs do not depend on the distance between two locations. However, Mansori (2003) is an exception in the literature by introducing increasing returns to scale in trade costs.
[7] In our conclusions we explain some weakness of the NEG approach.

Krugman (1979) introduces for the first time a model of monopolistic competition with international trade. This paper is the genesis of the NT literature and is a natural reference to NEG models. Its setting generates intra-industry trade between countries with identical technology and endowments. Indeed, the comparative advantage cost theory predicts no trade among countries with similar preferences, technology and factor endowment. This NT static model uses Dixit and Stiglitz (1977) framework by incorporating in his model one industrial sector with firms that exhibit increasing returns of scale, imperfect competition, endogenous firm´s location, trade costs and labor immobility between countries. Its most remarkable outcomes are related to the gains of trade. In particular, within a monopolistic competition framework and two countries without labor mobility between them, international trade openness implies that some firms are forced to exit and the ones that still remain in business expand out their production, and consequently, operating at lower average cost. In other words, the number of product varieties produced in one country decreases after trade barriers fall. The first source of gain from trade comes out of the love of variety principle: in each country consumers have access to more product varieties with both local and foreign origin. The second source of gain comes out higher wages. However, if preferences are represented by a CES function, then varieties produced do not vary in each country. In other words, there are no gains from taking advantage of the scale of production, and consumers gain solely by having more varieties.

An extension of Krugman (1979) is Krugman (1980), which is also a seminal work on NT theory that formalizes the concept of "home market effect "(HME) and "the price index effect" (PIE). It underlies the importance of initial market size to determine the national industrial structure. It takes the first four ingredients listed in our introduction. In each country two types of differentiated goods are produced; each one has its own consumers. The proportion of the population in country *i* that consumes good *j* is defined as the market size of good *j*. Labor in this sector is fixed, immobile and equal across countries. Thus, we have two countries with equal population, technology but trading both types of good. Different markets sizes across countries is allowed. There are four important outcomes which at some extend keep being valid in further models. One is the HME: the country with the larger market size of a particular good will attract disproportionately more firms that produce such a good, and therefore become a net exporter. Two, incomplete specialization is greater, the greater trade costs and the less important scale economies are. For example, with very low trade costs and large economies of scale countries will thoroughly specialize. Three, incomplete specialization implies that each country will export all of its varieties of both types of goods. And finally four, is the PIE that arises when consumers share the same utility function: the larger a country is, the lower industrial price index have because a small

proportion of this country's industrial good consumption bears trade costs. Yet, Krugman (1980) is unable to explain that small shocks can lead to permanent effects.

Helpman and Krugman (1985) is a generalization of Krugman (1980). The innovations in this framework are three. First, there are two sectors, a commodity that is produced under constant returns to scale in a perfect competitive market and traded at zero cost; and an industrial good that is a differentiated good in a monopolistic competition market. The second element is that population (labor force) across countries is not equal. And finally, preferences are Dixit and Stiglitz (1977) type. The HME and the PIE keep being valid.

Agglomeration in Krugman (1980) and Helpman and Krugman (1985) can only arise through the magnification of the initial market size asymmetries. Nevertheless, Davis (1998) modifies slightly Helpman and Krugman´s (1985) assumptions to turn down the HME. In particular, it is assumed that the commodity is traded at a positive cost. The industrial's good assumptions on technology and market structure keep being untouched. This tiny departure, which is empirically justified, gives the following proposition. Each country will produce the industrial good variety according to each country's market size, and the agricultural good will not be traded whatsoever. In sum, location effects vanish. This is how the model works. Assuming that each country produces the commodity good according to its own requirements hence in each country industry sector is also distributed according to its market size. If some firms move into the larger country, then trade in the commodity good increases, whereas trade in the industrial god falls. Given that trade costs are equal for both goods, then total aggregated costs of trade increase and shifted firms find the move unprofitable.

In sum, in NT models large regions firms will be net exporters with higher relative wages. For Ottaviano and Puga (1998), this approach still has important shortcomings that are attacked by the NEG approach. First, NT theory conceives differences in production structure through differences in underlying characteristics. It starts by assuming that there are regions with large and small markets, but does nothing to explain why this division arises. Second, it does not explain why firms in particular sectors tend to locate close to each other, leading regional specialization .Third, it presents industrial development as taking place gradually and simultaneously in all underdeveloped regions, while in practice industry spreads successively from country to country.

NEG is set up, among others, by Krugman (1991) core-periphery (C-P) model, which is based upon Helpman and Krugman (1985), shows that small temporary shocks give rise to large permanent differences between two regions. One is the core of industrial production and the other is a periphery, which employs all of its labor force in the commodity production. The new ingredient

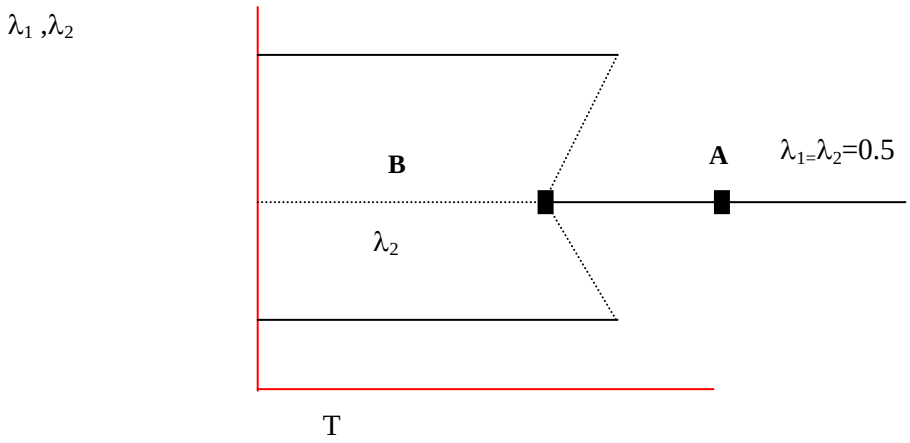
in this framework is factor mobility: labor force can decide the location to carry out its activities. It turns out that the HME could be exacerbated by the combination of increasing returns to scale and imperfect competition. It consists of two regions which are identical in endowments, technology and preferences. In each region, there is a labor endowment consisting of farmers and workers. There are two sectors in each region, commodity and industrial. The former exhibits constant returns to scale technology with farmers as the only factor of production. It produces a homogenous good sold in an interregional competitive market. The commodity sector trade costs are neglected. The latter sector technology exhibits increasing returns to scale with workers as the only factor of production. The industrial good is sold in an interregional imperfect competitive market, where there firms produce a different variety of the industrial good and exit and entry is costless. The manufacturing sector trade costs are Samuelson type and could be positive. Only workers can move across regions. Both farmers and workers have a common Cobb-Douglas utility function with preferences over the commodity and a CES utility function, which incorporates *n* differentiated products.

A stable and dispersed equilibrium arises for prohibitive trade costs. It consists of two identical economies in autarchy: wages, prices, output and varieties are determined within each region. No trade takes place. A core-periphery stable pattern, in which the whole industrial sector is agglomerated in one region, arises in the following way. If a larger number of firms is located in one region (a deviation from the dispersed equilibrium) a circular causation is generated through forward and backward linkages (linkages between firms and workers/consumers). More firms imply more variety of products, and lower prices and profits. Such a situation attracts more workers from the other region due to higher real wages (forward linkages). More consumers implies a larger demand and ease competition in the labor market that attracts more firms (backward linkages). More firms implies more variety of products, and lower prices and profits. This agglomeration process emerges if trade costs fall below a critical level. Therefore, through these linkages effects, scale economies at the individual firm level are transformed into increasing returns at the level of the region as a whole. In this case a stronger market competition associated with more firms is dominated by location decisions of firms. This dynamics depends on historical accidents: small initial differences trigger this evolution process.

We cannot fail to notice the following exotic dynamics of Krugman´s (1991) predictions (see figure 2). First, we find a non-negative relationship between trade openness and concentration; however, the shape of the stable equilibrium is discontinuous and non monotonic. Second, a gradual fall of trade costs does not imply anything, in terms of stability, except in some specific range of

trade costs where a deviation could arise an abrupt equilibrium change. And third, with low trade costs full agglomeration is predicted. Mossay (2006) proves the existence and uniqueness of the short-run equilibrium of Krugman´s (1991) C-P model.

Figure 2. Krugman´s Fundamental Relationship between Trade Costs and Agglomeration



3. The Bell-Shaped Curve of International Trade Openness

How trade openness in the form of a bilateral trade arrangement or a multi-country union custom, may change industrial location and wages around the world? Accordant with Krugman and Venables (1995), over time policy circles have had two opposite perspectives over the impact of globalization on the North-South divide. On the one hand, during the 1950s, 1960s and 1970s they claimed that integration produced a rise in the living standards of rich nations at expense of the poor ones. Accordingly most of the developing countries implemented trade policies that followed the "import-substitution industrialization" paradigm, which supports the idea of low levels of international openness as an optimal policy to foster internal industrialization.[8] For example, Krugman and Hanson (1993) point out that Mexico in the last century undertook protectionist trade policies to avoid a dependent relationship with the U.S. economy. Krueger (1997) explains that developing countries were also motivated to close their markets based upon the infant industry argument: new firms face higher costs relative to incumbent firms operating abroad. On the other hand, Krugman and Venables (1995) claim that during the 1990s there was a growing concern in the developed countries on the effects of integration. In the U.S. "respectable voices" considered that local jobs would move to Mexico searching for lower wages and more flexible regulations. In

[8] However, some countries like Hong Kong, Taiwan, Korea and Singapore shifted toward outward-oriented policies.

Europe, official documents claimed that developing manufacturing countries had an adverse impact on its employment rates.

What explains this reversal in the conventional thinking? In a world with two identical countries, in terms of tastes and technology, Puga (1999) theoretically reconciles both visions and displays a different menu of possibilities than standard trade theory does. At intermediate trade costs industrial location has a C-P pattern. However, as trade barriers fall industrial concentration gradually vanishes. Furthermore, at zero trade costs welfare convergence is also reached between these two countries, which is also a result in Krugman and Venables (1995). In sum, the curve that shapes the share of industrial location or welfare as trade costs fall looks like a bell à la Kuznets. Puga (1999) is inspired by a very important question regarding European integration: Will European economic geography features, like income disparities across regions and manufacturing concentration, converge to that of the U.S.? At regional level, where labor mobility is allowed, Wheaton and Shishido (1981) also reconcile both visions by arguing that a clear dominance of the prime city and a widening urban-rural wage gap are highly expected to come up during early stages of economic growth. As development proceeds, spatial dispersion and narrowing wage differential should occur. Hence, the emergence of a C-P pattern would be followed by convergence.

Krugman and Venables (1995)[9] is a seminal paper that formalizes the bell-shaped curve of economic change. It is the international version of Krugman´s (1991) C–P framework. Two new assumptions are incorporated. First, it rules out regional labor mobility but incorporates labor mobility across sectors. Put another way, the labor agglomeration mechanism is domestic, so when a sector expands the labor supply must come from the other sector. Wages in the other country is not a dispersion force anymore. Second, the industrial sector uses part of its own production as inputs. This assumption creates new cumulative agglomeration forces known as forward and backward linkages. Both forces arise when firms simultaneously consider the other firms as suppliers and consumers of inputs, respectively. Recall that Krugman´s (1991) model centrifugal forces decline with trade costs at an even more rapid rate than the centripetal forces that promote agglomeration

Their main results can be divided into three parts. In the first one, trade costs are prohibitive, then a symmetric and stable equilibrium arises. In this equilibrium both regions are characterized by zero profits, equal real wages across sectors, same price for each variety and positive activity in both sectors. Any deviation from this outcome, for example, when the number of manufacturing firms increases in one region, affects firm's profitability through four channels. The standard

[9] The working title for Krugman and Venables (1995) is "History of the World, Part I".

channel (á la Chamberlain) reduces the profits by shifting down the demand that each firm faces. However, the channel called forward linkage reduces total and marginal costs because inputs are cheaper. The backward linkage shifts the demand up of each firm because the total expenditure on manufactured products also increases. Both linkages generate higher profits. The stability of this equilibrium rests on the net outcome generated from this deviation. Finally, the labor market channel increases wages costs due to a higher local labor demand. The negative effect on profits of the standard and the labor markets channels outweighs the forward and backward linkages effects.

The second part of this story starts when trade costs fall below a critical threshold. Both symmetric and asymmetric equilibria, which are stable are possible.[10] In the asymmetric equilibrium, the world arises into a high real wage industrial "core" and a low real wage agricultural "periphery". In the core region the price index is low and nominal wages are equal or greater than one, thus real wages are high and all labor force is concentrated in the manufacturing sector. Consumers import all their agricultural goods and import a small amount of manufactured goods. In the periphery region the price index is high and nominal wages equal to one, thus real wages are low and most labor force is concentrated on the agricultural sector. Most of the manufactured goods are imported.[11]

The third part comes up for lower trade costs, where only the asymmetric equilibrium is sustainable. As the transport costs keep declining real wages in both regions converge in a non monotonic way, in particular they describe a bell-shaped pattern in the core region. The lower transportation costs are, the weaker the forward and backward linkages in the periphery region are, thus firms star moving to the periphery region because wages are lower. At zero transportation costs real wages are higher than real wages in the symmetric equilibrium with prohibitive transportation costs. It is worth mentioning that Krugman and Venables (1995) focus their attention on welfare implications of trade openness rather than industrial clustering.

Venables (1996) provides some notions of another agglomeration force through backward and forward linkages, which are already present in Krugman and Venables (1995). Even without labor mobility an input-output structure may constitute a force of agglomeration. It assumes two regions and three sectors. The commodity sector´s technology exhibits non-increasing returns to scale, whereas the other two industrial sectors´ technology exhibit increasing returns and are vertically linked through an input-output structure. Downstream firms use an aggregate of upstream varieties as an intermediate output. Such a structure creates two agglomeration forces. One is a forward linkage which push upstream sector to increases their sales by locating where there are relatively

[10] For intermediate transport costs there are other two unstable equilibria
[11] It is possible to have a extreme C-P pattern at some level of trade costs.

many downstream firms. The other one is a backward linkage which pushes firms in the downstream sector to reduce costs by locating where there are relatively many upstream firms. The fact that both upstream and downstream industries are monopolistically competitive makes the agglomeration forces arise solely through market interactions. By assuming interregional labor immobility the location of the demand works as an opposite force of agglomeration. The balance of these centripetal and centrifugal forces depends on the strength of the vertical linkages and trade costs.

Economic integration that implies lower trade costs will lead to either divergence or convergence between regions. The final outcome depends on the strength of both vertical and trade costs. For weak vertical linkages and low trade costs, then firms' location depends on wage differences and dispersion is a feasible outcome. For strong vertical linkages and intermediate trade costs clustering may arise. Another conclusion is related to welfare implications of industrial clustering. Firms clustering together attract more firms and can support a relative high wage. A clue element in this model is that imperfect competition allows that vertical linkages get a relevant role.

Puga and Venables (1997) is a generalization of Krugman and Venables (1995) for *M* countries. The authors cope with the location effects of geographically discriminatory trade policy. More precisely, they analyze welfare implications of economic integration by considering three cases: Global integration, free trade areas and hub-and-spoke arrangements. Their key assumption is that in the manufacturing sector firms require final goods as inputs. Under global integration, all firms regardless of their location have equal access to any foreign market. For high trade costs, each country is self-sufficient, with production domestically oriented in both sectors. A symmetric equilibrium arises where all nations have the identical values for all endogenous variables. If the trade costs fall below a threshold, an asymmetric equilibrium arises where its precise characterization varies with the number of nations and the share of industry in consumer expenditure. When there are two nations we return to Krugman and Venables (1995).

The second case is related to preferential arrangements like NAFTA: Trade openness takes place in a club of two or more countries but each member implement independent trade policies with the rest of the world. If they share their trade policy they become a custom union like the E.U. or MERCOSUR. For *M=3,* where two countries move toward a free trade area and the third one is outside the club the following immediate consequence arises: The number of firms increases and welfare in each country that belongs to the free trade area and decreases in the third one. The intuition behind this result is that firms within the free trade area face lower costs compared to firms outside the area. Thus, firms are attracted to countries that belong to the free area. As integrations

proceeds the countries within the area converge in welfare but not in industrial share. The country outside the area is negatively affected in its welfare and industrial share.

Finally, hub-and spoke arrangements are bilateral trade agreements between a country (the hub) and a set of countries (the spokes); however, among the latter ones there are trade barriers between them. A case is the association agreements between E.U. and some Eastern European countries. For M=3, where one country has a trade agreement with the other two countries, but these ones haven't liberalized their trade among them. The immediate results are that the number of firms and welfare increases in all countries, however, the change is larger for the hub than for the spokes. As integration proceeds welfare converges but not completely.

Puga (1999) is a major contribution to the NEG literature. As result of the interaction between the agricultural sector and the manufacturing sector in an international context, the exotic dynamics of location and trade costs relationship is eliminated. Recall that in Krugman (1991) factors are specific to each sector and in Krugman and Venables (1995) the labor's supply elasticity from the agricultural sector to the manufacturing sector is perfect. In both cases, agglomeration does not affect wages in the agricultural industry. Puga (1999) two novel assumptions are that we have decreasing returns in agriculture and firm entry and exit is a gradual process:

The first case is when wage differentials are eliminated by allowing interregional mobility in a context of input/output linkages as Krugman and Venables (1995) and Venables (1996) model. Labor distribution across sectors is endogenous to the model. For high international trade costs (T_0), the symmetric equilibrium is stable. If we do not assume input/output linkages we return to Puga (1999); further assuming that the distribution of workers is exogenous we have Krugman (1991) framework. If $\tau_A<T_o$, then we have a unique symmetric and stable equilibrium. In this case, if one region had more firms than the other, then competition will be stronger and profits would turn profits negative, inducing firms to relocate in the region with fewer firms. If $\tau_A>T_o>\tau_S$, then the symmetric equilibrium is still stable but is no unique; there are two stable agglomeration equilibria. In this case, full agglomeration, say region 1, is possible because input/output linkages are strong enough and trade costs are low such that is possible to compete in distant markets. It is worth mentioning that profits for any deviant firm to region 2 are negative ensuring stability of the equilibrium. If $\tau_S > T_0$ then the symmetric equilibrium is unstable but is no unique and the two agglomeration equilibria keep being stable. Any deviation from the symmetric equilibrium raises profits in region with more firms and reduces profits with fewer firms, then industry will eventually agglomerate.

The second case or the international version does not allow regional labor mobility, then labor endowment is fixed in each region and real wages are not required to be equal across regions in equilibrium. At high trade costs, firms locate according to the market size. At intermediate trade costs firms locate according to backward and forward linkages. At low trade costs firms locate where wages are lower. The contribution of this model is that it gets rid of the discontinuity of the share of the industry curve.

In figure 3, $\lambda f(h)$ denotes the fraction of the population in the foreign (home) country in the industrial sector. In figure 4, $rwf(h)$ denotes the real wage of the population in the foreign (home) country in the industrial sector.

Figure 3. Puga´s Bell Shape of International Trade Openness

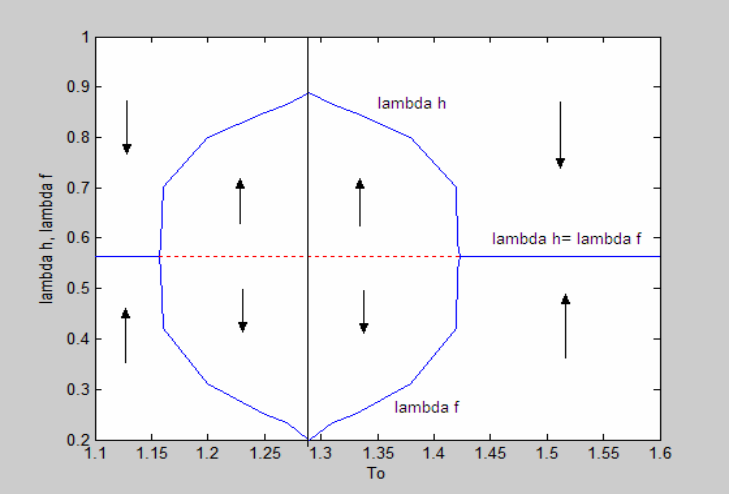
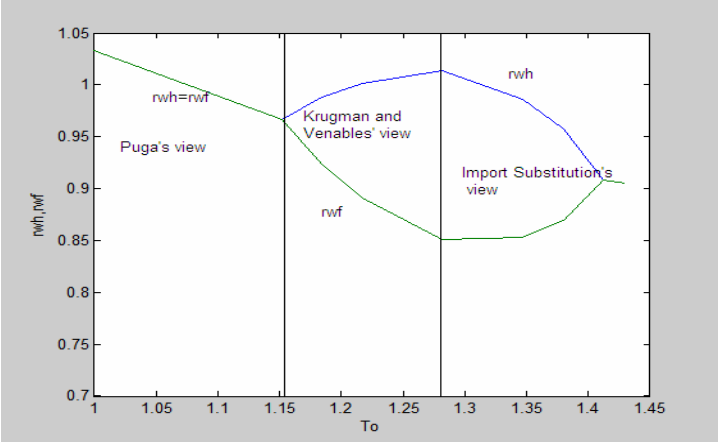


Figure 4. Puga´s Bell Shape of International Trade Openness



So far trade openness has been considered reciprocal between two or more countries. Puga and Venables (1999) address location effects of unilateral changes in trade policy by one active country. The first case is an import substitution policy, which successfully attracts industry. Under this policy there are two opposite effects. One is that as a result of higher prices of inputs incentives to firms to set up in the active country are weakened. But pulling in the opposite direction is the increasing in expenditure on industrial goods in the active country. The second case is trade liberalization also promotes industrialization in the active country. Within an interval trade costs induce zero industrialization. Above that range the active country attracts firms but real income has not so evident increase. Below this interval real income is higher the lower trade costs are and attraction of firms takes place as well.

Empirical Literature

Forslid et al. (2002) apply a full scale computable GE-model to investigate whether the outcomes and rationale of stylized NEG models, like Krugman (1991) and Krugman and Venables (1995), keep being valid in a more complex world. Traditionally, NEG models simplify their settings by dealing with two locations, two industries and two factors economy structure. This paper is based on Haaland and Norman (1992) model with the following departures. They assumes 10 regions (4 are associated with Europe: North, South, West and Central), 14 sectors and 3 factors of production, and both intra-industry and inter-industry linkages. For each predetermined level of trade costs the full set of parameters are obtained by three ways: calibration, assumption and secondary sources. In this framework, the relative weight of concentration forces depends on the level of trade costs. For high trade costs consumer proximity considerations determine location of production. For intermediate trade costs, input demand and input supply proximity considerations dominate location decisions. For low trade costs factor-market competition considerations determine location: specialization arises according to comparative advantage.

Their analysis proceeds in two parts. First, they show how production in different sectors changes as trade costs are reduced between the four European regions. Second, the authors simulate the absolute concentration index of the four European regions as trade costs are lowered.

In the first part the most striking result comes from the textile, leather and food sectors, which show a monotonic increase in agglomeration. Textile industry moves out of Central into West and South. Textile sector is a candidate for relocation effects because it has relatively strong within-

industry linkages. Initial textile production is slightly higher in West than Central. Then such a difference explains why production of textiles moves out of Central into West. Besides, South has a comparative advantage in the production of labor intensive goods as textiles. The leather sector concentrates exclusively in South. This can be explained by using the same arguments of the textile case. However, the initial production in South is considerably larger than the other regions. Besides as trade costs get lower the relocation movement is softer because it has a very low own input share. The food industry, which is relatively capital intensive, agglomerates in North for low trade costs. This can be explained by the comparative advantage of North. In addition to this, market size proximity is irrelevant because food industry is characterized by low returns to scale and a low own input share.

In the second part they simulate the location effects on industry at aggregated level in Europe. Textiles, leather products and food products concentrate in Europe with respect to the rest of the world as trade barriers fall; while metals, machinery and chemicals decreases. In the former case, a combination of comparative advantage factors and vertical linkages explain such movements. The latter is explained basically by increasing returns to scale.

Combes and Lafourcade (2004) evaluate the relevance of concentration and dispersion forces contemplated in NEG models for France. In particular, the authors estimate the parameters of an inter-regional trade model that includes two novel features. First, there are real strategic interactions and competition consequences within a Cournot framework. Secondly, labor market is neglected; in other words, there are no wage gaps. Two basic forces intervene in the location of firms. On the one hand, final and intermediate demands, and input costs are agglomeration forces. On the other hand, higher competition on the product markets is a dispersion force. They find that the centre (Paris) and its periphery (Marseille) firms' mark-ups are higher than middle points (Lyon). In the former case low trade costs offsets competition; in the latter case lower competition outweighs high trade costs. Furthermore, the economy displays a mono-centric pattern where Paris has larger profits and decay with distance. Lower trade costs reduce inter-regional disparities and intra-regional disparities.

According to Head and Mayer (2003) one of the research lines within NEG empirical work consider the impact of geographical distribution of demand as an explanatory variable. In this vein, Overman et al. (2003) find that variations in per capita income can be explained by the access to markets and sources of supply. In a first stage, trade equation, which is based on country dummies, provides estimates of market access and supply capacities. The authors then proceed in a second stage as follows. First they regress per capita income on market access controlling for other

determinants of income level like technology, resource endowments, other features of physical geography and institutional variables. Their main results are that market access is statistically significant to explain GDP per capita across countries. Second they regress machinery and equipment relative prices on supply access and find a negative relationship. Put another way, the better the supply access the lower are the inputs and factor production. And finally, predetermining the values of costs shares of intermediates and the elasticity of substitution between varieties, GDP per capita is regressed on both market and supply access. Overman et al. (2003) use an alternative trade equation, which is based on both country dummies, geographic and economic variables (access to coast, island status or distance to the E.U., U.S and Japan). Under this specification results have similar pattern of results. In addition to this, five countries are taken to predict changes in their GDP per capita as other geographic features change as well. For example, changing the status of landlocked countries like Paraguay or Zimbabwe increases substantially their GDP. Same thing happens by shortening the distance to Central Europe. The main conclusions are that wages do not determine location of firms. Other factors like geographical advantage are also significant for location. On other hand, for a given location of production distance keep being an obstacle for investment and trade. However, geographical advantages can improve as new industrial centers emerge.

Overman et al. (2003) decompose the South East Asian exports rate of growth into the contributions of improvements external demand and increased external supply. From a new trade model a system of equation is solved to obtain both foreign market and supply capacities growth rates for four periods. For 13 South East Asian economies exports performance have been remarkable. For example, Vietnam exports rate of growth from 82/85 period to 94/97 period was of 1512.52. Consequently, both market and supply access present remarkable rates of growth. The authors obtain the contribution of 9 regions to the South East Asian 13 countries.

By partially following, Head and Mayer's (2003) suggestions, Gatica Arreola and Ramirez Grajeda (2006) test Puga's (1999) fundamental bell-shaped relationship between trade openness and agglomeration in the industrial sector. In a world with two countries, they estimate the theoretical range of international trade costs in which agglomeration is expected: the share of industry, in terms of production or employment, is larger than its labor endowment share. On the other hand, from bilateral trade and production data they obtain a theoretical level of trade openness. Therefore, their hypothesis according to Puga (1999), states that the shorter is the distance from this value to the middle point of the interval, the larger is agglomeration. With information on 28 OECD countries, 14 years and 29 industrial branches, they find that for every sector, the employment and production

gap gets larger as the level of trade openness gets closer to the center of the agglomeration interval. Nevertheless, there is no empirical support pertaining to the impact on the employment share.

4. Metropolis and International Trade

According to Alonso-Villar (2001), urban centers were an exceptional phenomenon until the 19th century. For example, classical Rome in the 1st century, clusters of business in the Middle Ages like Venice or Bruges, or capitals of the new absolutist states in the 17th and 18th centuries like London or Paris. Political centralization and the growth of international trade led to the overall urban population localized in only few cities, all of which were capitals. Nevertheless, Ades and Glaeser (1995) argue that the growth of Tokyo (Edo) in the 18th century is explained by Krugman and Livas (1996) international trade hypothesis. A centralized regime reduced foreign trade, which strengthened the centripetal forces of agglomeration that fostered an urban "giant"[12].

For Alonso-Villar (2001), modern urbanization started during the Industrial Revolution in cities of the United Kingdom like Birmingham, Leeds or Manchester, which attracted new labor force. This process extended across other countries like Germany, northern France and the east coast of the U.S.. Before 1900 this process was pretty much European. In traditional societies the functions of cities were mainly administrative, commercial, religious and craft-related. Yet during the 20th century concentration of population has appeared not only in Europe but also around the world. In the last decades urban population in Latin America, Asia and Africa has grown dramatically. For Henderson (2001), 75 per cent of Latin America's population is urbanized and 30 per cent in Asia.

Inspired by the case of Mexico City, Krugman and Livas (1996) argue that Third World metropolis will tend to shrink as developing countries open their markets. Trade openness within a country involves larger markets for any of its production sites, driving firms to relocate close to foreign markets such as border regions or port cities. Incentives to move out are stronger for small countries because its local market represents a low proportion with respect to its foreign markets. Other papers as Venables (1998), Alonso-Villar (2001) and Mansori (2003) address the link between trade openness and spatial considerations as well.

Krugman and Livas (1996) consider that there are centripetal and centrifugal forces whose balance depend on trade costs and determine industry agglomeration. Centripetal forces involve, in Hirschman´s (1958) words, backward and forward linkages. The former are related to market

[12] In the late 16th century, Japan was unified by Tokugawa Ieyasu who concentrated high levels of economic and political power. Ieyasu descendants closed the country to any foreign contact. In the mid 19th century, Japan was forced by the US to open its economy to foreign trade.

access; the latter are related to good access to intermediate inputs. Centrifugal forces are external diseconomies, land rents and the attraction of moving away from highly competitive urban areas to less competitive rural ones. They focus their attention on the Mexico City case where the centripetal forces traditionally have dominated the centrifugal forces. Mexico was a closed economy under the Import-Substitution Industrialization paradigm. However, once Mexico was opened up to international markets, domestic final goods demand and domestic input supply weight less as a centripetal force. The existence giant Third World metropolis are a consequence of strong backward and forward linkages. Policies which tend to open the economy weaken these linkages and, consequently, foster dispersion.

Krugman and Livas (1996) formalize their story through a mathematical model. In this survey we present an extended model featured in Fujita *et al.*(1999). There are four cities which are thin and narrow. Cities 1 to 3 are domestic locations and city 0 is considered the rest of the world. The only factor of production is labor, which is fixed and can move across domestic cities. Within each city real wages net of a congestion cost are equal across agents. If there is a difference in wages between cities 1 and 2 people start moving to the city where the wages are higher. Agglomeration makes sense because the existing technology exhibits increasing returns to scale.

There are two assumptions in Krugman and Livas (1996) to preserve the constant elasticity of demand facing firms. One is the usual iceberg type trade costs of goods between local cities of $1/T$.; and two, an iceberg type international trade cost of $1/T_0$ for imported goods from location 0, which results from a combination of transportation costs and trade protectionism. Both the cost for people of moving from one domestic city to the other and exports costs are zero. Although Krugman and Livas´ (1996) model is quite simple, it is too complicated to be solved analytically. So they present a numerical example. And using the tricks of Dixit and Stiglitz (1977) they can get fundamental equations to explain the existence of big metropolis.

In both figures, the center represents equal distribution of the population across cities. Points (0, 0), (0.5, 0.86) and (1, 0) mean that the whole domestic population is concentrated in cities 1, 2 or 3, respectively. The middle point between the line that joins points (0, 0) and (0.5, 0.86) means that total population is equally divided between cities 1 and 2. In these figures, the initial point of an arrow is a point which represents a short-term equilibrium given a particular distribution of the population. This means that real wages might be different across cities, then labor immigration is expected to generate a new distribution. The length of the arrow represents the magnitude of labor movements over time across cities ($\Delta\lambda 1$, $\Delta\lambda 2$, $\Delta\lambda 3$) and the direction represents the sign of these changes ($\Delta\lambda j \geq 0$ or $\Delta\lambda j < 0$).

Figure 5 shows that for high levels of international trade costs (T0 = 1.9), partial concentration in one city is a stable long-run equilibrium. It should be pointed out that concentration in one city is not total because a small fraction of the total population is distributed across the rest of the cities. Equal distribution between three or two cities implies an unstable equilibrium. Internal and international trade takes place and all varieties produced in the economy are consumed in all cities. The main city produces a large variety of goods and the secondary cities produce a limited variety of goods and trade between cities is balanced. Figure 6 shows that the equal distribution of population in the domestic country is a stable long-run equilibrium for high levels of trade openness. Partial concentration in one or two cities is unstable.

With high international trade costs, both firms and workers, by emphasizing their expenditure on national goods magnify the market size effects of agglomeration through prices and nominal wages. In other words, an extra worker in a particular city represents a higher demand and such a benefit always offsets fiercer competition in the labor market. Thus, equilibrium is reached when congestion costs are high enough to prevent further agglomeration. For lower trade costs (T0 = 1) imports weight in agents' expenditure is large enough such that any deviation from the dispersed equilibrium is associated with weak market size effects.

The intuition behind Krugman and Livas´ (1996) results can be summarized as follows. This model suggests a link between protectionism and the size of big metropolis of protective countries. International firms supply every location in the country. Domestic firms pay lower transport costs when serving their own location. Then, domestic prices, net of travel, are lower where domestic firms are agglomerated. Trade barriers imply that domestic suppliers take over the market. Prices, net of transport costs, are lower for domestic goods in the central city because firms are located in that city. Workers then come to the city to pay lower prices for domestic goods. Trade openness implies that imported goods are a large part of consumption. Imports are more expensive in the central city, so workers spread over space to save on congestion costs.

Figure 5. Urban Agglomeration without International Trade

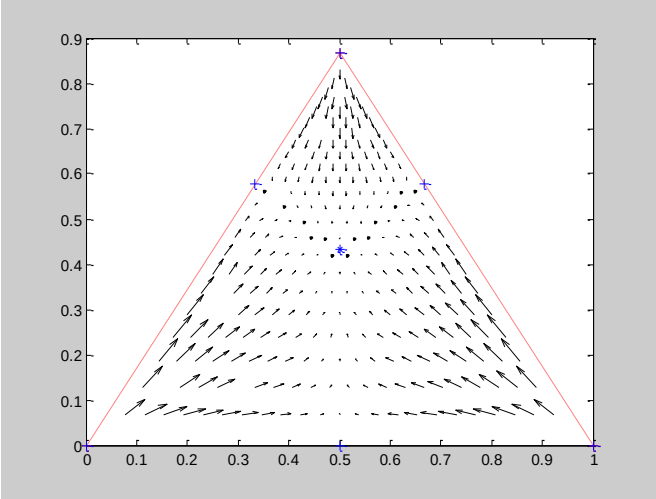
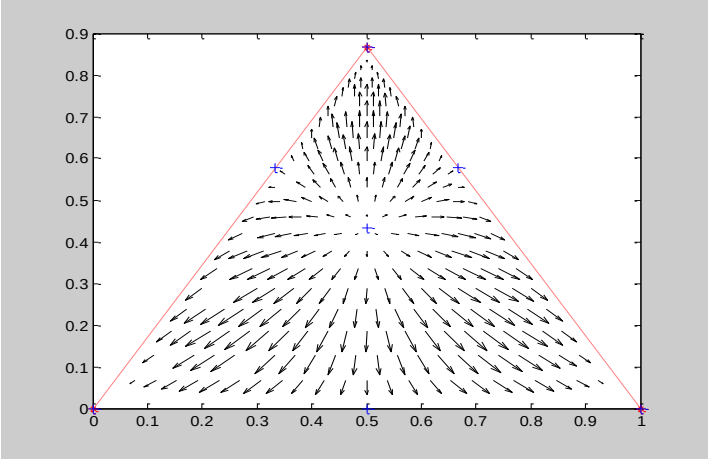
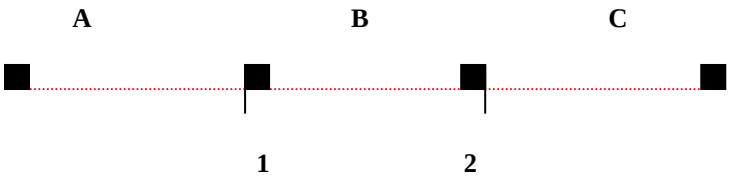


Figure 6. Urban Agglomeration with International Trade



Alonso-Villar (2001) follows the Krugman and Livas´ (1996) setting by arguing that agglomeration arises as a result of increasing returns of scale, transports costs, labor mobility and the relative position of a country with respect in terms of industrialization. Congestion is the selected dispersion force. The model assumes an economy as a horizontal line (see Figure 7) divided into three segments: A, B and C. The left segment, A, has a city in its extreme. The middle segment, B, has two cities, 1 and 2, in each extreme. The right segment, C, has a city in its extreme.

Figure 7. Alonso Villar´s Urban Structure



World population is normalized to 1. λi is the proportion of world population in location i (i=a, 1, 2 and c). This paper analyzes the centripetal and centrifugal forces, which drive concentration in one of the cities in segment B. First, the paper analyzes the factors that affects agglomeration. Alonso-Villar (2001) defines the short term equilibrium as Krugman (1991) does: given an initial distribution of the population between locations the model determines prices, amounts of goods, number of firms and wages in each city given labor immobility. The long run equilibrium is divided into two cases, autarchy and free trade. Under autarky, the results are that under high congestion costs, any wage differential between city 1 and 2 take λ1 back to the original point λ1=0.5; under low congestion costs the original equilibrium is unstable. Therefore, agglomeration is less likely for high congestion costs. Under free trade, if λa=λc, λ1=λ2 an λb is high then the original equilibrium is stable. It means that trade openness does not weaken the original agglomeration forces. If λb is low then the original equilibrium is unstable if trade openness is above a threshold. In sum, if the Dominican Republic has a BTA with the U.S., it will concentrate its population in the capital if there is a wage gap with respect to the second most important city. The U.S. spatial organization will be untouched.

Contrary to Krugman and Livas (1996) and Alonso-Villar´s (2001) results, Mansori (2003) concludes that under increasing returns to scale in the cost of trade, trade liberalization may cause big cities to concentrate even more industry. His assumption is that trade of costs are positive for local and foreign transactions. However, there are two types of outcomes after trade barriers fall. One is that some megalopolis that are already in equilibrium do not change their size; the other is that the size can be larger. Buenos Aires and Bangkok fall in the former case; the latter could be Seoul. Mansori (2003) has four conclusions. First, in welfare terms a dispersed equilibrium is preferable that a C-P pattern. Second, infrastructure improvements can deviate a country from C-P pattern equilibrium to a dispersed one. Third, a country can move from a dispersed equilibrium to a C-P pattern as a result of trade openness. And finally, trade openness can negatively affect welfare because gains from trade can be offset by congestion costs that arise from concentration.

Empirical Literature

Krugman (1998b) considers that empirical work has failed to validate theoretical models of monopolistic competition. The new industrial organization has been notoriously better a creating interesting models that at generating empirical predictions. The new growth theory gave rise to a massive industry of cross-country growth regressions, but with few exceptions, these regressions have been neither closely tied to the theory nor a clear evidence (Recall: Sala-i- Martin ran two million regressions!). In this section, we present two remarkable papers in the literature. Ades and Glaeser (1995) and Hanson (1998).

Ades and Glaeser (1995) investigate the forces that drives the concentration of a nation's urban population in a single city. They define two types of forces. Economic forces as high tariffs, low levels of international trade and high costs of internal trade. First, high tariffs, high costs of internal trade and low levels of international trade increase the degree of concentration. Second, political stability negatively affects urban population's share. Third, they conclude that political factors affect urban concentration but not the other way around. In part, validates Krugman and Livas (1996) approach.

They use a sample of 85 observations (the main city in 85 countries). Their main results are the following. First, main cities are 42 percent larger in population, on average, if they are also capital cities. This fact means that power attracts population and that capitals are located in larger cities. Second, a 10 percent increase in the size of the country increases population in the main city by about 1.2 percent. Third, a one standard-deviation increase in the share of trade in GDP reduces the size of the main city 13 percent. Fourth, main cities are 45 percent larger in countries with dictatorial regimes. And fifth, a rise 1 percent increase in the ratio of import duties rises the size of the central city by 3 percent.

Hanson (1998) summarizes the literature on changes in spatial organization among North American countries after NAFTA. After forty years of industrialization based on the import-substitution paradigm, Mexico opened its economy to international trade in 1986 by becoming a member of the GATT. Hanson (1996, 1997) find that trade liberalization has contributed toward the breakup of the traditional manufacturing belt on Mexico City, and the formation of new industrial centers in the US-Mexican border. Hanson (1996) finds that with trade liberalization, there was a substantial relocation of manufacturing activity in the US-Mexican border. There is a significant relation of export firms in Mexico and economic activity in the U.S. border. U.S. cities specialize in products and components for Mexican assembly plants. Besides, NAFTA has push firms from the interior of U.S. to U.S. border cities. Hanson (1997) finds a negative relation between wages and

the distance to Mexico City and distance to Mexico-U.S. border. A 10 percent increase in distance from Mexico City is associated to a 1.9 percent decrease in the relative state nominal wage. A 10 percent increase in distance from Mexico-U.S. border is associated to a 1.3 percent decrease in the relative state nominal wage. These results suggest that differential in market access foster wage differentials.

Davis and Weistein (2002) analyze the population distribution of Japan under several approaches including the increasing returns to scale theory. They use a 7,000 years database from the Stone Age to current times. They find that only the fact that density population variation raised during the industrial revolution is consistent with the increasing returns to scale theory, but persistence in regional densities and mean reversion after temporary shocks. Da Mata et al. (2005) find that city growth in Brazil is driven by rural population supply and inter-regional transport improvements and spillover effects of knowledge accumulation.

Ramirez Grajeda and Sheldon (2009) draw upon Fujita *et al*. (1999) as a theoretical motivation, and information on the 5 most important cities of 84 countries, to find that the size of main cities declines and the size of secondary cities increases as a result of external trade. Similar results are obtained for cities with a population over a million. However, cities with a large fraction of the urban population grow independently of their position in the urban ranking.

5. Concluding Remarks

In this survey we cope with trade, development and location issues under the New Economic Geography (NEG) approach. Despite the fact that this literature is relatively new exists a consensus within the economics profession that its main theoretical outcomes are very appealing. However, it is common knowledge that NEG predictions still need to be validated. This task is far from being easy for the following fundamental reason: since location issues imply increasing returns to scale, then non-linear relationships arise. Furthermore, some of the most representative papers lack of analytical solutions and their setting are highly stylized. As a result of this technical obstacle, empirical work is not abundant and robust enough. Additionally, lack of data prevails.

The relationship between space and international trade has come up in mainstream economics during the last years, since Krugman (1991). Yet empirical work is also scarce and has weak conclusions. Although location considerations have a long tradition, theoretical development is still young and it is covered by a limited number of economists. Several unrealistic assumptions in the standard literature seem worth pointing out in order to foresee future research. First, population is

exogenous; second, distance is generally neglected. Third, agents do not have expectations; Four, the analyses take locations as given. In sum, NEG simplify its models assumptions for tractability motives but that might limit its prediction power.

References

Ades, A. F. and Glaeser, E. L. "Trade and Circuses: Explaining Urban Giants." *The Quarterly Journal of Economics*, 1995, 110 (1), 195-227.

Alonso, W. *Location and Land Use*. Cambridge, MA: Harvard University Press, 1964.

Alonso-Villar, O. "Large Metropolises in the Third World: An Explanation." *Urban Studies*, 2001, 38 (8), 1359-1371.

Arthur, B. W. "Competing Technologies, Increasing Returns and Lock-in by Historical Events." *Economic Journal*, 1989, 99 (394), 116-131.

Brakman, S.; Garretsen, H. and Van Marrewijk, C. *An Introduction to Geographical Economics*. Cambridge: Cambridge University Press, 2001.

Brakman, S. and Garretsen, H. "Trade and Geography: Paul Krugman and the 2008 Nobel Prize for Economics." *Spatial Economic Analysis*, 2009, 4 (1), 5-23.

Chamberlain, E. *Theory of Monopolistic Competition.* Harvard Economic studies, 1933.

Christaller, W. *Central Place in Southern German*. in Jena Fisher, 1933.

Combes, P. P. and Lafourcade, M. "Trade Costs and Regional Disparities in a Model of Economic Geography: Structural Estimation and Prediction for France." Manuscript, 2004.

Da Mata, D; Deichman, U; Henderson, J; Lall, S. and Wang, H. "Determinants of City Growth in Brazil." NBER, Working Paper 11585, 2005.

Davis, D. R. "The Home Market, Trade and Industrial Structure." *American Economic Review*, 1998, 88 (5), 1264-1276.

Davis, D. and Weistein D. "Bones, Bombs and Break Points." *American Economic Review*, 2002, 92 (5), 1269-1289.

De la Blache, V. *Principes de Géographie Humaine*. Paris: Gallimard, 1921.

Dixit, A. and Stiglitz, J. "Monopolistic Competition and Optimum Product Diversity." *American economic Review*, 1977, 67 (3), 297-308.

Edwards, S. "Openness, Productivity, and Growth: What Do We Really Know?" *Economic Journal*, 1998, 108 (447), 383-398.

Forslid, R.; Haaland, J. I. and Midelfart Knarvik K. H."A U-Shaped Europe? A Simulation Study of Industrial Location." *Journal of International Economics*, 2002, 57 (2), 273-297.

Fujita, M. "Monopolistic Competition and Urban Systems." *European Economic Review*, 1993, 37 (2-3), 308-315.

Fujita, M. and Thisse, J. F. "Economics of Agglomeration." *Journal of the Japanese and International Economies*, 1996, 10 (1), 339-378.

Fujita, M. and Thisse, J. F. "New Economic Geography: An appraisal on the Occasion of Paul Krugman´s Nobel Prize in Economic Sciences." *Regional Science and Urban Economics*, 2009, 39, 109-119.

Fujita, M.; Krugman, P. R. and Venables, A. *The Spatial Economics: Cities, Regions and international Trade*. Cambridge, MA: MIT Press, 1999.

Gatica Arreola, L. A and Ramirez Grajeda, M. "The Bell-Shaped Curve of International Trade Openness: A panel Data Test for OECD countries." *Econoquantum*, 2006, 3 (1) January, 113-133.

Hanson, G. H. "Integration and Location of Activities: Economic Integration, Intraindusty Trade, and Frontier Regions." *European Economic Review*, 1996, 40 (3-5), 941-949.

Hanson, G. H. "Localization Economies, Vertical Organization and Trade." *American Economic Review*, 1996, 86 (5), 1266-1278.

Hanson, G. H. "Increasing Returns, Trade and the Regional Structure of Wages." *The Economic Journal*, 1997, 107 (440) January, 113-133.

Hanson, G. H. "North American Economic Integration and Industry Location." *Oxford Review of Economic Policy*, 1998, 14 (2), 30-43.

Harris, C. "The Market as a Factor in the Localization of Industry in the United States." *Annals of the Association of American Geographers*, 1954, 64, 315-348.

Harrison, A. "Openness and Growth: A Time Series, Cross Country Analysis for Developing Countries," *Journal of Development Economics*, 1996, 48 (2), 419-447.

Head, **K. and Mayer, T. "**The Empirics of Agglomeration and Trade" Manuscript, 2003.

Helpman, E. and Krugman, P. R. *Market Structure and Foreign Trade: Incresing Returns, Imperfect Competition, and the International Economy*. Cambridge, MA: MIT Press, 1985.

Henderson, V. J.; Shalizi Z. and Venables, A. "Geography and Development." *Journal of Economic Geography*, 2001, 1 (1), 81-105.

Hirschman, A., *The Strategy of Economic Development*. New Haven, CT: Yale University Press, 1958.

Hotelling, H. "Stability in Competition." *Economic Journal*, 1929, 39 (153), 41-57.

Krueger, A. "Trade policy and economic development: how we learn", *American Economic Review*, 1997, 87 (1), 1-22.

Krugman, P. R. "Increasing Returns, Monopolistic Competition and International Trade." *American Economic Review*, 1979, 69 (4), 469-79.

Krugman, P. R. "Scale Economics, Product Differentiation and the Pattern of Trade." *American Economic Review*, 1980, 70 (5), 950-59.

Krugman, P. R. "Increasing Returns and Economic Geography." *Journal of Political Economy*, 1991, 99 (3), 137-150.

Krugman, P. R. "What's New about The New Economic Geography." *Oxford Review of Economic Policy*, 1998a, 14 (2), 7-17.

Krugman, P. R. "Space: The Final Frontier." *Journal of Economic Perspectives*, 1998b, 12 (2), 161-174.

Krugman, P. R. and Gordon, H. "Mexico-U.S. Free Trade and The Location of Production," in Garber, P., ed., *The Mexico-U.S. Free Trade Agreement*. Cambridge, MA: MIT Press, 1993.

Krugman, P. R. and Livas Elizondo, R. "Trade Policy and The Third World Metropolis." *Journal of Development Economics*, 1996, 49 (1), 137-150.

Krugman, P. R. and Venables, A. J. "Globalization and the Inequality of Nations." *Quarterly Journal of Economics*, 1995, 110 (4), 857-880.

Limao, N. and Venables, A. J. "Infrastructure, Geographical Disadvantage, Transport Costs, and Trade." *The World Bank Economic Review*, 2001, 15 (3), 451-479.

Lösch, A. *The Economics of Location*. 1954, New Haven, CT: Yale University Press.

Mansori, K. S. "The Geographic Effects of Trade Liberalization with Increasing Returns in Transportation." *Journal of Regional Science*, 2003, 43 (2), 249-268.

Marshall, A. *Principles of Economics*, London. Macmillan ,1920.

Mossay, P. "The core periphery model: a note on the existence and uniqueness of short-run equlibrium." *Journal of Urban Economics*, 2006, 59 (3), 389-393.

Ottaviano, **G. I. and Thisse, J. F.** "On Economic Geography in Economic Theory: Increasing Returns and Pecuniary Externalities." *Journal of Economic Geography*, 1997, 1 (2), 153-179.

Ottaviano, **G. I. and Puga, D.** "Agglomeration in the Global Economy: A Survey of the New Economy Geography." *The World Economy*, 1998, 21(6) 707-731.

Ottaviano, **G. I. and Thisse, J. F. "**Agglomeration and Economic Geography." Manuscript, 2003.

Ottaviano, **G. I. and Thisse, J. F. "**New Economic Geography: What about N?" Manuscript, 2004.

Overman, H.; Redding, S. and Venables, A. "The Economic Geography of Trade, Production and Income: A Survey of Empirics," in Kwan, C. and Harrigan, J. eds. , *Handbook of International Trade*, Blackwell Publishing, 2003.

Pred, A. *The Spatial Dynamics of US Urban-Industrial Growth, 1800-1914*. Cambridge, MA: MIT Press, 1966.

Puga, D. "The Rise and Fall of Economic Inequalities." *European Economic Review*, 1999, 43 (2), 303-334.

Puga, D. and Venables, A. J. "Preferential Trading Arrangements and Industrial Location." *Journal of International Economics*, 1997, 43 (3-4), 347-368.

Puga, D. and Venables, A. J. "Agglomeration and Economic Development: Import Substitution vs. Trade Liberalisation." *Economic Journal*, 1999, 109 (455), 292-311.

Quigley, J. M. "Urban Diversity and Economic Growth." *Journal of Economic Perspectives*, 1998, 12 (2), 127-138.

Ramirez Grajeda, M. and Sheldon, I. M "Trade Openness and City Interaction." manuscript, 2009.

Samuelson, P. A. "The Transfer Problem and Transport Costs: The Terms of Trade when Impediments are Absent." *Economic Journal*, 1954, 62 (254), 278-304.

Scitovsky, T. "Two concepts of External Economies." *Journal of Political Economy*, 1954, 62 (2), 143-151.

Starret, D. "Market Allocations of Location Choice in a Model with Free Mobility," *Journal of Economic Theory*, 1978, 17 (1), 21-37.

Tirole, J. *The Theory of Industrial Organization*. Cambridge, MA: MIT Press, 1988.

Venables, A. J. "Economic Integration and the Location of Firms." *American Economic Review*, 1995, 85 (2), 296-300.

Venables, A. J. "Equilibrium Locations of Vertically Linked Industries." *International Economic Review*, 1996, 37 (2), 341-359.

Venables, A. J. "The Assessment: Trade and Location." *Oxford Review of Economic Policy*, 1998, 14 (2), 1-6.

Von Thünen, J. *Der IsolierteStaat in Beiziehung au Landschaft und Nationalökonomie*. Hamburg, 1826.

Weber, A. *Theory of Location of Industries*. Chicago, Il: University of Chicago Press, 1909.

Wheaton, W. and Shishido, H. "Urban Concentration, Agglomeration Economies, and the Level of Economic Development." *Economic Development and Cultural Change*, 1981, 30 (1), 17-30.

Yanikkaya, H. "Trade Openness and Economic Growth: a cross Country Empirical Investigation," *Journal of Development Economics*, 2003, 72 (1), 57-89.