CARTA ECONÓMICA REGIONAL | NUEVA ÉPOCA | ISSN 0187-7674 | AÑO 21 | NÚM. 103 | SEPTIEMBRE-DICIEMBRE 2009 | pp. 7-26

Spatial implications of international trade under the new economic geography approach*

Adrián de León Arias and Mauricio Ramírez Grajeda

Abstract In 2008, Paul Krugman, professor in 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 theoretical literature, known as the New Economic Geography (NEG), launched by Krugman (1991). In particular, we focus on four topics: *(a)* NEG roots, *(b)* NEG rationale; *(c)* the spatial impact of international trade on global economic imbalances; and *(d)* the impact of international trade on urban structure.

KEYWORDS New Economic Geography, trade oppenness, agglomeration and urban economics.

Resumen En 2008, Paul Krugman, profesor de la Universidad de Princeton, ganó el premio Nobel de Economía por su análisis teórico sobre «los patrones del comercio internacional y la localización de la actividad industrial». En este artículo revisamos la literatura de la Nueva Geografía Económica (NGE) iniciada por Krugman (1991). En particular, nos enfocamos en cuatro tópicos: a) las raíces intelectuales de la NGE, b) la racionalidad de la NGE c) las implicaciones espaciales del comercio internacional en los desequilibrios globales, y d) el impacto del comercio internacional en la estructura urbana.

PALABRAS CLAVE Nueva Geografía Económica, apertura comercial, aglomeración y economía urbanas.

* Conacyt partially supported this investigation. We thank Paulina Brambila Trejo and Antonio Romero Luján for their assistance. This is an original paper. An extended version of this paper is available online at Munich Personal RePEc archive.

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). Another example, given by Venables (1995), is the concern about the spatial implications of the European Union (E.U.) enlargement by the end of 2004. 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.

In this regard, 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. 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).

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. It is worth mentioning some previous surveys which have also focused on the NEG background.[1] Brakman y Garretsen (2009) argues 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 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

1 Some of these surveys do not focus on NEG background entirely, however because its relevance we mentionted them here.

main contributions of location theory and standard economic theory to NEG. In this paper, 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.

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. Finally, some implications of our survey, in terms of main NEG shortcomings and the way forward, are presented.

Intellectual Underpinnings: The NEG roots.

While constant returns to scale imply that activities are divisible, and thus, each activity can be carried out at any scale without sacrificing efficiency. Autarky becames 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. Here, following Fujita *et al.* (1999), we identify the antecedents of NEG into three alternative theories on industrial location: Marshall-Scitovsky externalities, urban economics and regional science.

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

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

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.

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 is a single urban center. Outside the urban center, agricultural producers sell their crops in the city. There are positive trades costs associated with transporting agricultural goods 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) 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.[3]

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)[4] 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

3 We recommend Brueckner (2000) as an excellent introduction to urban sprawling.

4 Their work constitutes what is known as central-place theory.

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.

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 a different perspective, Myrdal (1957) offered an interesting explanation on the institutional causes of regional cumulative causation.

In Harris (1954) 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 tends 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. 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.

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), who we surveyed in the later section, 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 while there may be six kinds of centripetal forces. NEG picks up a particular centripetal force: the pecuniary component of the market size Marshallian externality. For Henderson *et al.* (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.

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, by solely taking the pure externality component of informational spillovers under perfect competition. 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.

Those 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 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.

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.[5] Under this type of costs the constant elasticity of demand is preserved.

NEG models rely heavily on ad hoc, although realistic assumptions. [6] Head and Mayer (2003) catalog five key ingredients of this paradigm: Scale economies at firm level, imperfect competition, trade costs, endogenous firms» locations and endogenous location of demand.

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.

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

5 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.

6 In our conclusions we explain some weakness of the NEG approach.

models. Its setting generates intra-industry trade between countries with identical technology and endowments. 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 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. There are four important outcomes which at some extend keep being valid in further models. One is the HME: countries 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. 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. Helpman and Krugman (1985) is a generalization of Krugman (1980) where 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. 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 imply 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 or: small initial differences trigger this evolution process (path dependency). We cannot fail to notice the following exotic dynamics of Krugman's (1991) predictions (see figure 1). 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.

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.[7] 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.

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)[8] 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.

Their main results can be divided into three parts. In the first one, trade costs are prohibitive

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

8 The working title for Krugman and Venables (1995) is «History of the World, Part I».

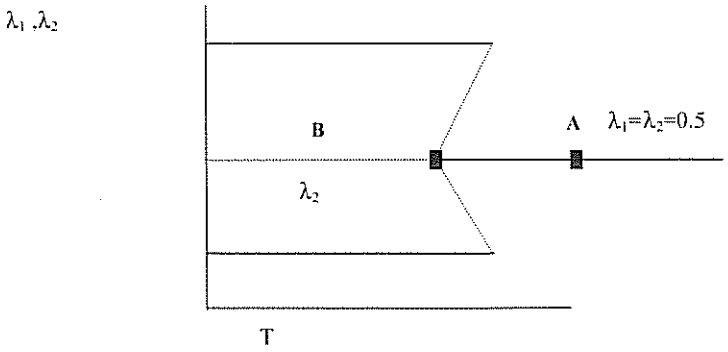

FIGURE 1. Krugman´s Fundamental Relationship between Trade Costs and Agglomeration.

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 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.[9] 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.[10]

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

9 For intermediate transport costs there are other two unstable equilibria

10 It is possible to have a extreme C-P pattern at some level of trade costs.

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 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 key 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. 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).

In the second case 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_o), 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 $\tau_S > T_o$ 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 2, where we illustrated the Puga 's Bell Shape of International Trade Openness, $\lambda f(l_i)$ denotes the fraction of the population in the foreign (home) country in the industrial sector.

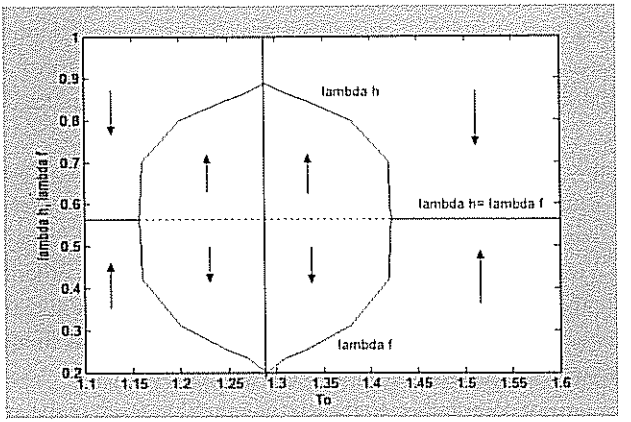

FIGURE 2. Puga's Bell Shape of International Trade Openness.

Up to this point 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.

Metropolises and International Trade

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. De León (2003 y 2004) evaluate empirically the implications of the Krugman and Livas´ hypothesis for the Mexican case.

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

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.

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 figures 3, 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 3 shows that for high levels of international trade costs ($T_0 = 1.9$), 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. However, if the distribution of population is equal in the domestic country, then a stable long-run equilibrium arises for high levels of trade openness. Partial concentration in one or two cities is unstable.

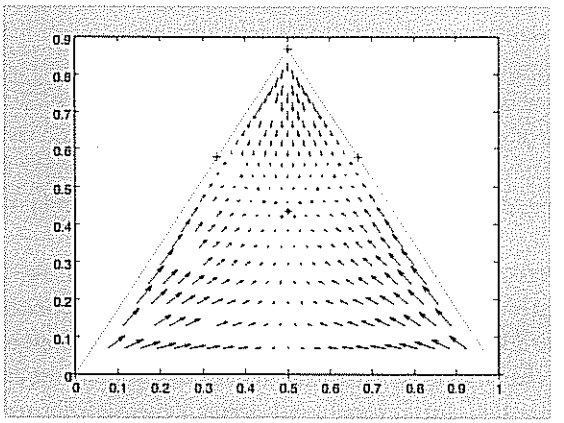

FIGURE 3. Urban Agglomeration without International Trade.

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 ($T_{\rho} = 1$) 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.

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. She adds a new foreign country and suggests that the negative relationship between trade openness and city size depends on the relative size of the home country. If it is small with regard to the rest of the world, a dispersed equilibrium is not sustainable given low levels of trade costs.

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 lo-

cal and foreign transactions. Mansori (2003) has four conclusions. First, in welfare terms a dispersed equilibrium is preferable that a C-P pattern. Second, infrastructure improvements can shift 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.

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. And finally, exogenous initial conditions determine the long-run equilibrium. In sum, NEG simplify its models assumptions for tractability motives but that might limit its power of prediction.

References

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

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

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

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

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

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

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

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

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

De León Arias, A. (2003). «Trade Liberalization and Growth: Evidence from Mexican Cities.» *The International Trade Journal*, XVII (3).

De León Arias, A. (2004). «Trade Liberalization and Productivity Growth: Some Lessons from the Mexican Case.» *Investigaciones Geográficas, Boletín del Instituto de Geografía de la* UNAM.

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

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

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

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

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

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

Head, K. and Mayer, T. (2003). «The Empirics of Agglomeration and Trade» Manuscript.

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

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

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

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

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

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

Krugman, P. R. (1981). «Intra-industry Specialization and the Gains from Trade» *Journal of Political Economy*, 89: 959-73.

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

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

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

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

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

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

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

Marshall, A. (1920). *Principles of Economics*, London: Macmillan.

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

Myrdal, G. (1976). *Teoría Económica y Regiones Subdesarrolladas*, México: Fondo de Cultura Económica.

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

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

Ottaviano, G. I. and Thisse, J. F. (2003). «Agglomeration and Economic Geography». Manuscript.

Ottaviano, G. I. and Thisse, J. F. (2004). «New Economic Geography: What about N?» Manuscript.

Overman, H.; Redding, S. and Venables, A. (2003). «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.

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

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

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

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

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

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

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

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

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

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

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

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

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