New Economic Geography and Economic Integration: A Review

Andrea Ascani, Riccardo Crescenzi, Simona Iammarino

Department of Geography and Environment
London School of Economics and Political Science
Email: a.ascani@lse.ac.uk

Abstract
This review explores the main contributions to New Economic Geography (NEG) with a particular focus on the effects of economic integration on spatial development. Firstly, the theoretical framework is explored by presenting the fundamental building blocks of NEG and successively looking at the principal models of NEG. Then, empirical research within a NEG framework is summarised. Mostly, relevant empirical studies are surveyed to give a sense about the main paths that research has covered so far and, more particularly, the debate about the economic integration effect of the EU enlargement to Central Eastern European Countries (CEECs) is explored.

Keywords New Economic Geography, Economic Integration, Agglomeration, EU Enlargement

Acknowledgments:
This research was supported by the Project Sharing KnowledgE Assets: InteRregionally Cohesive NeigHborhoods (SEARCH) within the 7th European Community Framework Programme FP7-SSH - 2010.2.2-1 (266834) European Commission.
1. Introduction

As two or more countries integrate one major concern arises with respect to the effects that economic integration might have on the distribution of income between countries as well as on the welfare levels of regions within these countries. In traditional trade theory, when economic liberalisation occurs between two countries both of them benefit of the gains of comparative advantages (although these benefits may not be equally distributed). This fundamentally results in higher consumption levels in each country as a result of trade as compared to a situation of autarchy. However, while the notion of economic integration mainly evokes trade-related issues it also entails a wide number of relevant elements for economic geography and spatial development. Although it is only in recent years that these elements have been explored more deeply, it is not new story that the concept of economic integration, along with international trade theory in general, is intimately and unfailingly connected with location issues (Ohlin, 1933; Predöhl, 1950; Balassa, 1967). What is the impact of trade liberalisation on the geographical distribution of industries? How does economic integration shape spatial disparities? Why firms agglomerate in certain places as integration deepens? These and other questions crucially underpin the emergence of what is now well-known as New Economic Geography (hereinafter NEG), that is, a body of research initially stemming from international trade theory which fundamentally attempts “to explain the formation of a large variety of economic agglomeration (or concentration) in geographical space” (Fujita and Krugman, 2004, p. 140). Most of the concepts and tools employed by NEG as well as the ambiguous impact of economic integration on development were well-known before NEG appearance. For instance, the crucial role of increasing returns to scale for agglomeration to occur is anticipated by Myrdal (1957)’s concept of ‘cumulative causation’, the importance of externalities for localisation is firstly discussed by Marshall (1890) and the fact that economic integration might reveal detrimental effects for the economic performance of less developed regions to the advantage of those initially developed is anticipated by Kaldor (1970). Yet, the innovative contribution of NEG consists of the rigorous formalization of such concepts which basically allows accounting for the dynamics of spatial clustering (and dispersing) of economic activity when trade barriers are progressively removed, which is hardly explainable with traditional theory.

The aim of this paper is to review the main contributions to NEG with a particular focus on the effects of economic integration on spatial development. Firstly, the theoretical framework is explored by presenting the fundamental building blocks of NEG and successively looking at the principal models of NEG. Then, empirical research within a NEG framework is briefly summarised. Due to the asymmetry between theoretical and empirical NEG in terms of contributions and relevance, this paper pays more attention to the former. Mostly, relevant empirical studies are
surveyed to give a sense about the main paths that research has covered so far and, more particularly, the debate about the economic integration effect of the EU enlargement to Central Eastern European Countries (CEECs) is explored. Clearly, the information included in this text is not exhaustive of the literature on NEG.

2. The building blocks of New Economic Geography

Mechanics of NEG is based on a number of fundamental elements that provide a plausible theorization of why self-reinforcing centripetal forces that pull economic activity into a location occur and persist over time. More particularly, increasing returns to scale, monopolistic competition, transaction costs and the occurrence of external economies collectively underpin the general functioning of NEG models and thus shape firms’ and workers’ location behaviour. As a result, the combination of such theoretical tools and the occurrence of specific parameters values in the economies modelled by NEG make it possible to explain the geographical unevenness of the economic landscape as a situation of equilibrium.

Firstly, increasing returns to scale are acknowledged to be fundamental when accounting for the spatial unevenness of economic activity given that they allow considering geography as a fundamental element in the analysis. In fact, as Scotchmer and Thisse (1992, p. 272) highlight, the importance of increasing returns constitutes the so-called ‘folk theorem of spatial economics’ since they by definition stimulate economic production to cluster in space. Indeed, as NEG models allow for increasing returns to occur, manufacturing firms are strongly encouraged to concentrate production in space as a way to benefit from the advantages of scale economies. In other words, increasing returns represent a notable incentive for firms to geographically concentrate their productive activities rather than dispersing them in several locations due to the benefits in terms of production costs deriving from creating larger plants. In this respect, increasing returns crucially constitute a sort of leitmotiv of NEG which is central to the explanation of the spatial differences in the distribution of productive activities. However, the mere existence of increasing returns does not imply that production is automatically concentrated in space. In fact, the agglomerating effect of increasing returns is the complex result of the interaction with other forces within the economy, as explained in the remainder.

Secondly, monopolistic competition, based on works such as Chamberlin (1933), Spence (1976), Lancaster (1979) and in particular Dixit and Stiglitz (1977) enters NEG as a decisive element that

---

1 Relevant theoretical reviews of NEG include Ottaviano and Puga (1998) and Baldwin et al. (2003) whilst Overman, Redding and Venables (2003), Head and Mayer (2004b) and Redding (2010) represent notable empirical surveys.
underpins the existence of scale economies in formal models. Indeed, including economies of scale implies that competition between firms is far from perfect since each firm can increase production while reducing the average cost per unit of product (Samuelson and Nordhaus, 2001). In contrast, in perfectly competitive markets the assumption of increasing returns (internal to the firm) cannot hold as the cost of producing an additional unit of product necessarily implies negative profits. Moreover, the existence of increasing returns allows for the creation of larger plants that are in turn more efficient than smaller ones since when a firm decide to concentrate production in one single location the benefits of scale economies give it an advantage over spatially dispersed firms. This is extremely different from a situation of perfectly competitive markets where constant or decreasing returns eliminate the occurrence of economies of scale internal to the firm. In fact, in this framework firms are not concerned with any location choices since they cannot benefit from increasing returns by concentrating production. Thus, they will decide to produce in all locations where consumers are, thus distributing economic activity as an optimum. Hence, adopting imperfect competition in the economic reasoning becomes essential for considering the benefit of scale economies and explaining spatial pattern in the location of economic production. Generally, monopolistic competition includes horizontally differentiated products with constant elasticity of substitution (CES) so that consumers buy small amounts of each different variety of the good (i.e. varieties enter demand symmetrically). Therefore, each firm operating under increasing returns produces just a single variety of the differentiated good and thus decides its price. As a result, each firm tends to operate monopolistically in the specific market related to the variety that it produces. Since the size of each market is restricted by the existence of alternative varieties of the differentiated product, the monopolistic power of each firm over price is limited by the presence of other firms. Similarly to perfectly competitive markets, monopolistic competition exhibits a wide number of producers as well as the absence of barriers to entry or exit the market. Generally, when operating in a monopolistic competition setting, firms are neither in perfect competition nor in a situation of monopoly (Combes et al., 2008). By adopting such an arrangement NEG basically traces market and demand structure while dealing with increasing returns, thus providing a theoretical framework that permits to investigate the formation of economic agglomeration in space.

Thirdly, transport costs are included in NEG as a crucial element that influences location choices. Whereas in most traditional trade theory such costs equal zero by assumption, NEG generally adopts some forms of ‘iceberg transport costs’ à la Samuelson (1952) where only a fraction of the value of the units of product shipped from a location to another arrives while the rest is paid as cost of shipment. Therefore, the impact of transport costs on firms’ location choices clearly depends on the level of such costs. As a consequence, firms decide whether it is more convenient to concentrate in just a single location and serve other regions by exports or alternatively incur in additional fixed costs to open up a second plant in a different location. With this in mind, the interaction between the
level of transport costs and increasing returns constitutes a crucial force towards agglomeration (or dispersion) in firms’ location behaviour.

Finally, external economies are incorporated in NEG to give an account for the high level of localisation of individual industries or, for analytical purposes, of localisation of the manufacturing sector as a whole (see Krugman, 1991a, p.485). In doing this, NEG essentially recalls Alfred Marshall’s insights about externalities: labour market pooling, availability of specialised intermediates and technological spillover effects. Firstly, firms that cluster in a single location take advantage of the availability of pooled labour force endowed with industry-specific skills. On the workers’ side, clustering firms represent a situation in which the risk of unemployment is reduced as compared to an economy where firms are dispersed. In general, there is an increase in efficiency emerging from an agglomerating industry connected with a local pooled labour market. As Krugman (1991b) argues, the benefits of market pooling are realised only in presence of increasing returns to scale which actively encourage firms to locate into a single location. Secondly, when firms concentrate production into a single location they also take advantage of the presence of specialised suppliers of intermediate goods and inputs. This means that through the creation of backward and forward linkages between producers of final goods and their suppliers of intermediates a self-reinforcing efficiency gain is determined at the industry level (see Krugman and Venables, 1995). In other words, firms tend to concentrate into a single location to take advantage of scale economies creating an incentive for suppliers of intermediates to locate production in the same location. In turn, as production of final goods by clustered firms becomes gradually less expensive due to better access to intermediates, more firms are attracted into the same location and this effect reinforces industry concentration. Evidently, the efficiency gain of such an external economy is triggered by the presence of increasing returns in the production of both final and intermediate goods. If this is not the case, “even a small-scale center of production could replicate a large one in miniature and still achieve the same level of efficiency” (Krugman, 1991b, p.49). Thirdly, clustered firms are supposed to benefit from technological spillovers consisting in unintentional flows of knowledge arising from proximity to one another and benefitting the industry as a whole. As a result, firms are encouraged to localise in a single place to benefit from external knowledge arising from other firms’ activities (i.e. R&D). However, while acknowledging the relevance of technological spillovers in explaining the localization patterns of firms, NEG does not focus much on such an externality since it fundamentally presents some operationalization difficulties. Indeed, NEG authors rather prefer to deal with pecuniary externalities for which demand and supply can be individuated and modelled. In this respect, both labour market externalities and firms’ forward and backward linkages can be concretely accounted for whereas technological spillover effects are much more uncertain and invisible (Krugman, 1991a and 1991b). As mentioned above, a central point in the theorisation of external economies in NEG is that their effect arises only in presence of increasing returns internal to
firms which are then encouraged to concentrate production into a single location as a way to reduce costs.

3. Market access and the emergence of a core-periphery pattern

Krugman and Venables (1990) provide one of the first relevant formal contributions about the functioning of NEG models in predicting agglomeration in a framework of economic integration. Their source of inspiration is the European 1992 Single Market and the previous Southern enlargement which made them investigating the effects of the interaction between market access and integration on the competitiveness of the manufacturing sector in the new access countries. Basically, Krugman and Venables (1990) model a two-sector and two-country economy where the existence of increasing returns in the imperfectly competitive manufacturing sector encourages firms to concentrate production in a few places. The two countries exhibit different market access with the ‘larger’ country benefitting of a central position with better market access than the ‘smaller’ country occupying only a peripheral location with a lower access to demand. The level of trade costs enters the model in a crucial way since its variation ultimately influences the location decisions of firms. Agglomeration forces prevail for intermediate levels of barriers to trade, when proximity to the larger market attracts firms to locate in the core country. This allows firms not only to access a larger number of consumers without incurring in shipment costs, but also to serve the periphery through exports. In fact, for intermediate trade costs firms in the core can easily increase exports towards the periphery thus reducing the dispersion effect of competition in the larger market. In this scenario, peripheral consumers can be served by exports from the core and most firms relocate in the core country. Differently, in the case of high trade costs each country has a share of manufacturing that equals its endowments. Product market competition represents in this case a sufficient dispersion force towards a more even distribution of economic activity in space. In fact, if one of the two regions, say the centre, has more manufacturing firms than its market size would allow (leaving the periphery with less firms relative to its market size), then increasing product market competition and the lack of opportunities to increase exports due to high trade costs forces some firms to relocate from the centre to the periphery. Finally, for low trade costs location of manufacturing firms is mainly driven by factor market competition which acts as a dispersion force. Indeed, since exporting from the periphery to the core becomes less costly due to reduced trade costs more firms tend to relocate in the periphery as a way take advantage of factor price differentials between the two countries. Briefly, Krugman and Venables (1990) provide a helpful explanation of the ambiguous effect that economic integration has on the competitiveness of industry in the periphery.

4. The centripetal pull of labour mobility

6
One of the main shortcomings of Krugman and Venables (1990) is that they do not explain the process of emergence of differences in the production and market structure of the two countries considered. Indeed, the core or periphery status is attributed exogenously since countries are endowed with a large or small market as they enter the model. In other words, the authors do not answer the question “why countries that are *a priori* very similar can develop very different production structures” (Ottaviano and Puga, 1998, p.712). This appears to be a very relevant issue when considering the effects of economic integration on the location choices of firms since differentials in market access as in Krugman and Venables (1990) are just part of the story of why firms agglomerate. In this respect, Krugman (1991a) provides an endogenous explanation of the process underlying the occurrence of a core-periphery pattern in the case of two initially identical regional economies. The general setting still encompasses a two-sector and two-region economy where manufacturing activities operates under increasing returns to scale and the market for these goods is monopolistically competitive. The main difference with Krugman and Venables (1990) consists of allowing for labour migration from one region to the other as a response to market signals. This is a crucial difference, since the interaction of interregional migration with increasing returns and trade costs determines the balance between agglomeration and dispersion forces in the economy. When agglomeration forces dominate the resulting concentration process triggers a mechanism of ‘circular causation’ *à la* Myrdal which basically constitutes a self-reinforcing dynamics in the economy. In other words, when trade costs are sufficiently low, firms decide to locate where demand is larger in order to benefit from economies of scale, and demand becomes larger (i.e. more immigrants) as production of manufactures concentrate. This process feed on itself automatically, resulting in the emergence of persistent differences in the economic structure of the industrial core as compared to the agricultural periphery.

Three main forces shape the process of agglomeration/dispersion of economic activity in space. Firstly, the 'product market competition' effect implies that when one worker migrates from Region B to Region A competition in the latter raises (while it is reduced in the former). Then, firms pay lower wages in Region A relative to Region B as a way to support their competitiveness. This effect clearly constitutes a dispersion force since some workers in Region A will decide to migrate in Region B where the relative wage is higher. Secondly, the ‘home market effect’ implies that, other things being equal, the region with the larger market for a specific product has the higher wage and it is a net exporter of that product (Krugman, 1980): in fact, more workers in Region A entail a larger share of income spent in industrial goods and this allows local firms to pay higher nominal wages, making this location increasingly attractive for more workers (and consequently more firms). As such, Region A becomes an exporter of industrial goods. Thirdly, the ‘price index effect’ implies that a larger share of workers in Region A determines lower prices for industrial products in the local market. In fact, more varieties are produced in Region A and they do not incur in trade costs since
most firms produce locally. Thus, prices are lower in Region A relative to Region B. As such, the real wage in Region A as compared to real wage in Region B rises attracting more workers in Region A. The intensity of these three forces as well as the balance between them is determined by the level of trade costs between the two regions.

Let us start with a situation where trade costs are high: firms’ location behaviour is mainly driven by the competition effect that basically prevents the process of agglomeration and supports a situation where economic production is evenly distributed across space. Since each region has the same endowments (i.e. no a priori differences between regions), firms have no incentive to relocate from one region to another since they would face more competition without the possibility to serve the other region’s market by exports due to high trade costs. Hence, in this case, final demand is met locally and the share of manufactures in both regions remains the same. This is a stable equilibrium since, for high trade costs, dispersion forces prevail over agglomeration forces.

In order to illustrate this particular case of high trade costs, consider for example that for an exogenous reason (i.e. historical accident) one worker migrates from region B to region A. As a consequence, the latter has a larger share of manufacturing labour force than former and real wage ratio \( \frac{w_A}{w_B} \) between the two regions decreases. Firms in Region A face lower profitability due to more intense product market competition (i.e. more varieties produced) on the local market and high trade costs that impede exports. Hence, firms in Region A are forced to reduce wages as a way to increase their competitiveness on the local market. As a result, real wages tend to be lower in Region A. At the same time, firms in Region B face less competition on their local market and have higher profits. In this situation, manufacturing labour force in Region A start migrating to Region B until the ratio \( \frac{w_A}{w_B} \) equals 1, that is, until wage differential is eliminated and firms face the same degree of competition in both regions. Therefore, regional convergence is expected and the distribution of industry is basically shaped by that of agriculture, which represents a relevant part of demand for manufactures. In this case, agglomeration forces are not strong enough to prevail on the dispersion effect because firms in Region A cannot compete in distant markets (due to high trade costs). Workers have no incentive to migrate in Region A and agglomeration forces are not triggered. Contrarily, exceeding workers migrate back to Region B where relative wage is higher. As explained, this migration flow ends when wages are the same in both regions. However, if we consider trade costs that are sufficiently low, the dispersion effect is not strong enough to impede concentration. In fact, when a worker relocates from Region B to Region A (again for exogenous reasons), agglomeration forces prevail. More workers are attracted in Region A because this offers higher wages and more varieties. As a result, the larger the share of workers in Region A, the larger the share of income spent in manufacturing goods (i.e. larger demand) in this location and, thus, more firms are attracted from Region B to Region A in order to increase their profitability. The
competition effect is weak because low trade costs allow firms in Region A to serve distant markets in addition to the local demand. This dynamics determines a divergence in the economic structure of the two *a priori* identical regions and even with trade costs approaching zero this self-sustaining process does not reverse. The reason basically lies in the circularity of the process of agglomeration sustained by labour migration.

While migration across regions is allowed, workers remain bound to their sectors of origin in Krugman (1991a). Puga (1996) adopts a very similar framework to Krugman (1991a) where he considers the possibility for workers to move from one activity to the other. He suggests that agglomeration most likely occurs when the supply of labour is sufficiently elastic, so that firms can also draw labour force from the agricultural sector without notable increases in the rural wage rate. As such, sector migration from agriculture to manufacturing only slightly affects the wage differential between rural and industrial activities. In this case, agglomeration takes place since more rural workers find it convenient to move in industry where wages are relatively higher. On the contrary, if more workers relocating from agriculture to industry determine a disincentive for more workers to do the same because of the fall in manufacturing wages relative to agricultural wages, then agglomeration does not take place. This can be the case of inelastic labour supply from agriculture to manufacturing meaning that an initial inflow of rural workers into industry heavily affects the wage ratio between sectors.

5. Vertical linkages as drivers of firms’ co-location

The core-periphery patterns that emerge from the interaction of increasing returns and trade costs in the previous theoretical models is mainly based on market size considerations (Krugman and Venables, 1990) and labour mobility both across regions and sectors (Krugman, 1991a; Puga, 1996). However, as far as the EU is concerned, the mobility of workers does not really appear to play the role of an adjustment process to wage differential between countries, as migration in Europe is rather weak (Siebert, 1997; Obstfeld et al., 1998; Puga, 1996, 2002;). As a result, NEG models such as those considered above can only in part explain agglomeration processes in Europe since the differences in wages that start the self-reinforcing concentration of production through labour mobility do not seem to trigger the same dynamics in the European scenario. Patterns of agglomeration of economic activities in space are not only shaped by consumers’ final demand. In fact, a considerable part of the demand for manufacturing goods comes from other firms. In other words, firms producing intermediate goods represent a notable market for firms producing final goods. In this respect, Combes et al. (2008) estimate that the share of intermediate goods in the total manufacturing production of the US in 1997 equals 59%. That is, more than a half of total industrial output is consumed by other firms. Hence, this figure suggests that intermediate goods effectively
represent an important share of demand. Therefore, even in a setting where labour is immobile across regions agglomeration can equally take place through the interaction of increasing returns, trade costs and vertical linkages between firms.

In this respect, Venables (1996) constructs a NEG model where the main centripetal force for agglomeration arises from cost and demand linkages between firms. The manufacturing sector is split into a sector producing intermediate goods and another one producing final goods. Both of them operate under increasing returns and imperfect competition so that they have an incentive to cluster production in space in order to exploit economies of scale. Agriculture remains perfectly competitive and characterised by constant returns to scale. The idea is that firms producing intermediate goods decide to locate where the share of firms demanding intermediates is relatively higher. Therefore, in this framework, firms producing final goods create a demand linkage with suppliers. At the same time, firms consuming intermediate inputs will locate where the share of suppliers is relatively higher as a way to access goods by not incurring in trade costs. As such, also a cost linkage ties the two industries. Hence, being the location decisions of upstream and downstream firms mutually dependent and reinforcing, vertical linkages between firms represent a considerable force towards agglomeration. The existence of intermediate goods also implies the occurrence of other spatial forces. Indeed, given that labour is immobile across regions, when Region A has a relatively higher share of firms, the wage rate in this region increases as compared to Region B. As a consequence, demand of final goods grows in Region A attracting new firms from region B. This agglomeration process has a similar effect to the centripetal force seen in Krugman (1991a). Differently from Krugman (1991a), though, here it is an increase in income that affects final demand, rather than a migration-induced increase in local population. In addition, an increase in wages in Region A also triggers a dispersion force. In fact, firms in Region A face higher wages to pay relative to Region B. Therefore, in order to minimize costs firms may be induced to relocate where wages are lower. The balance of centripetal and centrifugal forces is mainly determined by the strength of linkages between firms as well as by trade costs. For high trade costs, the location decisions of firms are driven by market access considerations so that manufacturing of final goods is equally distributed between Region A and Region B. In fact, when trade barriers are high final consumer demand is mostly served locally. With increasing economic integration and trade costs approaching medium levels, differentials in costs between regions appear to be crucial. In this case, the region with more producers of inputs, say Region A, offers cost advantages for downstream firms, which start to move towards this location. As a result, the demand for intermediates in Region A becomes larger and more upstream firms are encouraged to relocate in such region, where demand linkages determine a larger volume of sales. Although rising wages in Region A constitute a force towards dispersion for firms, the interaction of scale economies with vertical linkages and intermediate trade costs makes centripetal pulls prevailing and industry agglomerates. In this case, then, the need of locating where
final demand is larger is less important than the need to locate where it is possible to exploit the advantages arising from the presence of firms in the other industry (i.e. downstream or upstream). For low trade costs, firms are dispersed across regions and the main driver of location decision tend to be the wage rate. In fact, inputs can be shipped without considerable transaction costs and vertical linkages are less relevant in location decisions.

Krugman and Venables (1995) offer a NEG model of vertical linkages which is very similar to Venables (1996). However, they consider just one manufacturing sector producing both intermediate and final goods. The interaction between trade costs and trade in intermediate goods encourages firms to cluster in order to exploit specific spatially-bounded pecuniary externalities (i.e. vertical linkages). The dynamics of such a model are generally the same of Venables (1996) with a non-monotonic relationship between the regional share of manufacturing and trade costs. A relevant contribution appears to be that of Krugman and Venables (1996). They consider a NEG model where vertical linkages are similar to Krugman and Venables (1995) with firms producing both final and intermediate goods. They explain the relationship between economic integration and industrial specialisation at the spatial level. In other words, such a study suggests that agglomeration processes may influence the location decisions of firms in the same industry leading to the emergence of specialised industrial districts. For high trade costs, each region will maintain an identical share of production in every industry, as usual in NEG models. As economic integration increases and the costs of trade approach intermediate levels, different results can occur. If industries are initially evenly distributed across regions agglomeration forces emerge but they are too weak to lead to a geographical concentration of firms. By contrast, in the case where the initial distribution of industries is uneven, centripetal forces are dominant at the sector level. In other words, if Region A initially exhibits a higher concentration of Industry K than Region B, then firms in Industry K located in Region B will find it more profitable to relocate in Region A to take advantages of more intense vertical linkages. Larger shares of firms in Industry K in Region A translate into larger local production of specialised intermediates, which in turn reinforce vertical linkages allowing firms to export at lower costs towards Region B. In addition, a larger industry in a region determines a relatively higher wage and this attracts new firms because consumers’ expenditure rises. Therefore, for intermediate trade costs, the initial distribution of industries appears fundamental for industrial specialisation to emerge. Finally, stronger economic integration leads to agglomeration of industries across regions because vertical linkages at the level of individual industries become crucial. In this case the agglomeration process takes the form of regional specialisation. This model appears to be particularly significant with respect to the European integration process, where a polycentric industrial geography exists mainly due to past barriers to trade as well as other differences between European nations (e.g. languages, cultures, etc.). Hence, this suggests that industrial specialisation patterns in Europe could occur more likely for very low trade costs. In fact, as noted by Krugman
and Venables (1996), intermediate economic integration could not be enough strong to lead production of specific industries to agglomerate in just one location as in the case of most industries in the US.

6. The interaction between labour mobility and vertical linkages

Puga (1999) provides a NEG framework combining the insights of models encompassing labour mobility across regions (Krugman, 1991a), across sectors (Puga 1996) and vertical linkages (Krugman and Venables, 1995). Four forces drive location decisions of firms between regions: product and labour market competition (centrifugal forces) and demand and cost linkages (centripetal forces). In this setting, interregional labour mobility fosters agglomeration since workers respond to wage differentials as in Krugman (1991a). As such, when for instance Region A offers a relatively higher wage than Region B, more workers locate in Region A, which in turn attracts more firms because of the larger demand. Contrarily, without migration, differences in wages across regions are not equilibrated by an inflow of labour force in Region A where the wage rate is higher. Therefore, some firms prefer to relocate from Region A to Region B in order to minimize production costs. This suggests that while labour mobility represents a force towards concentration of economic agents in a few places, the lack of interregional mobility, as in the European case, may delay agglomeration as economic integration proceeds. The crucial difference between considering or not labour mobility mainly refers to the case of low trade costs. In this case, labour mobility reinforces agglomeration given that the location of firms is primarily driven by final consumer demand. In fact, trade of intermediates can occur also over long distance without relevant additional costs and vertical linkages as an agglomeration force become weaker. However, when there is lack of labour mobility, low trade costs contribute to dispersion of production in space because some firms in the core move to the periphery where wage costs are lower; vertical linkages become less important due to the fact that strong economic integration allows to access intermediates by trade. Hence, the agglomeration pulls of linkages between firms weaken. In addition, factor market competition encourages firms to relocate where wage costs are lower, thus, dispersing manufacturing in space. For high and intermediate trade costs, instead, the agglomeration process is similar with or without migration across regions. When economic integration is weak firms decide to locate close to final consumer demand and markets are mainly served on a local basis. In this case, factor and product market competition discourage firms to relocate in the other region and vertical linkages are not strong enough to prevail on dispersion forces.

What is particularly interesting from this scenario with respect to Europe is that future European economic integration may benefit peripheral areas because of the low mobility of labour. The model highlights that the main dispersion force in this case is represented by factor market competition. In
fact, the occurrence of wage differentials between core and periphery encourages production to disperse for the benefit of peripheral areas. However, Puga (1999) notices that the existence in European countries of policies aimed at filling the wage differences at the subnational level is an element that weakens the effect of dispersion forces for high levels of economic integration. Thus, in the perspective of NEG, such political measures reduce the centrifugal effect of factor market competition in core areas by making wage costs in the periphery not advantageous for firms to relocate. Moving back to the theoretical model, for medium-level trade costs and labour mobility, agglomeration is favoured since economic integration allows firms to compete in distant markets and vertical linkages become a strong determinant of the concentration of production in space. In addition, even if vertical linkages are not considerable in magnitude, labour mobility still represents a relevant force towards agglomeration. With no labour mobility across regions agglomeration equally occurs. As a matter of fact, vertical linkages are a relevant force towards concentration, say in Region A. Furthermore, clustering firms in Region A start drawing labour from the local agricultural sector. As a result, industrial wages increase in Region A relatively to wages in Region B. This is possible because mobility across sectors is still an option in this framework and follows the same mechanisms of Puga (1996), where the high elasticity of labour supply from the rural to the urban sector allows industry to attract workers with only slight increases in rural wages, thus sustaining the flow of workers from agriculture to industry. Therefore, if firms choose to exploit local vertical linkages to access intermediates rather than importing such goods, then these demand and cost linkages à la Krugman and Venables (1995) represent pecuniary externalities that compensate for the higher wage rate that firms in Region A pay relatively to firms in Region B. As a result, agglomeration is sustained by vertical linkages and the inflow of rural workers in the urban sector. To conclude, when both vertical linkages and labour mobility are included in a NEG framework, the dispersion-agglomeration tension tends to follow a monotonic pattern as economic integration deepens. That is, as trade costs fall, economic activity simply concentrates in few places. Even for trade costs approaching to zero the process of concentration of production is not reversed. However, when labour is immobile across regions the agglomeration-dispersion relationship exhibits a non-monotonic trend for increasing levels of economic integration. In general, such a relationship between different levels of economic integration and the spatial distribution of industry seems to follow what is called the bell-shaped, or $\Omega$-shaped, curve of spatial development (Krugman and Venables, 1990, Puga, 2002; Combes et al., 2008); for high barriers to trade agglomeration is weak, for intermediate levels of economic integration strong agglomeration occurs and, finally, low trade costs reduce agglomeration. This suggests the fundamental ambiguity of the impact of economic integration on the occurrence of more or less stable core-periphery patterns in industry localisation.

7. **From theory to empirics: Applied NEG**
What we have seen above represents the main theoretical insights that NEG offers as far as economic integration and agglomeration processes are concerned. So far, NEG has mostly focused on theoretical models which allow making predictions about the effects of liberalisation on the location behaviour of economic agents. Empirical research within NEG appears much less developed relative to such an extensive body of theoretical work (Redding, 2010). A relatively small number of contributions analyse the economic integration impact of the EU enlargement eastward. This evidence is reviewed in this section right after a general overview of general empirical works using NEG tools. In the debate about the effect of integration on Central and Eastern European Countries (CEECs) NEG evidence is accompanied by some empirics from non-NEG literature which is nevertheless very interesting also in a NEG perspective due to the closeness of concepts employed as well as the insights offered.

7.1 NEG and economic integration

What most empirical studies using a NEG framework try to test is just the occurrence of one or more of the elements or forces that underpin agglomeration and dispersion in the economy. As we have discussed, market access represents a crucial driver of firms’ location decisions since choosing a region with larger market implies that trade costs are saved. Hanson (1996) represents one of the first attempts to test market access predictions of NEG in empirical work. He explores the effects of falling trade costs between Mexico and USA as a consequence of NAFTA on the location of Mexican manufactures. What Hanson (1996) suggests is that deeper economic integration has increased market access pulls for Mexican firms. In fact, most local production has relocated towards the bordering regions with the US. Moreover, integration has effectively transformed local firms from producers for the domestic market to product assembly for foreign-owned firms from the developed country. Therefore, the US-Mexican example highlights that economic integration between a developed and a developing area may influence both the geographical distribution of economic activities through market access considerations as well as the location of different stages of production across countries and regions. The importance of market access in the case of international economic integration is also highlighted by Overman and Winters (2006), who study the impact of the UK accession to the larger European market. They generally confirm NEG theoretical insights by explaining that regions hosting a port with better market access for exports and intermediate inputs experience higher employment rates. Instead, regions where accession has triggered a more intense product competition due to imports are characterised by a consequent decrease in employment. The importance of market access is also addressed by analysing the geography of factor prices. Redding and Venables (2004) and Hanson (2005) explore the relevance of spatial demand linkages at different geographical scales and both contributions suggest that wages vary spatially according to demand. In the same vein, Breinlich (2006) and Head and Mayer (2006) confirms the importance of proximity to large markets in shaping the core-periphery pattern of
regional per capita income in the EU. In Breinlich (2006), however, distance from larger demand seems to negatively impact wages in the periphery through scarce human and physical capital accumulation rather than through the occurrence of trade costs (see also Redding and Schott, 2003). Other empirical studies conducted for single European countries go in the same general direction confirming that a region market access is positively associated with higher wages (De Bruyne 2003; Mion, 2004; Brakman et al., 2004). Fallah et al. (2011) expand this line of research by exploring the distributional aspects of market access. Analysing US metropolitan areas, they suggest that not only wages are higher in areas with stronger market access, but also that wage inequalities between skilled and unskilled workers become larger in such areas. In fact, since better market access tends to be associated with economic sectors that are skill-intensive, increasing demand for skilled workers in these areas also determines a rise of their relative wage as compared to that of unskilled labour. Therefore, it emerges that the interaction between workers’ heterogeneity and market access may increase wage inequalities. In general, as suggested by most studies, demand linkages appear to be crucial for determining spatial patterns in the distribution of income and empirical research fundamentally confirms theoretical predictions.

Other authors have particularly focused on testing the occurrence and the relative importance of the home market effect that we have mentioned in previous sections. Davies and Weistein (1996; 1998; 2003) constitute the first main attempts to analyse the existence of such an effect. The aim of these studies is essentially that of investigating whether trade occurs as a result of traditional forces such as comparative advantage or because of increasing returns that give rise to the home market effect itself. While in Davies and Weinstein (1996) the NEG story of trade does not appear to be strong in explaining the structure of production in OECD countries, Davies and Weinstein (1998; 2003) are more refined studies that confirm that the home market effect is a substantial force. Also in other contributions, the home market effect is not always clearly individuated. For instance, Trionfetti (2001) highlights that the 'magnification effect' (i.e. the home market effect in author’s words) does not necessarily arise in every sectors of manufacturing activity. Head and Ries (2001) compare an increasing returns model of international trade à la Krugman with a model characterised by constant returns and find that in Canada and US a ‘reverse home market effect’ tend to dominate, i.e. an increase in the domestic demand for a specific product determines a reduction in the output of that product. However, other important contributions suggest that the home market effect exists and it is important. In addition to two later studies by Davies and Weinstein, Feenstra et al. (2001) study bilateral trade flows for Canada and suggest that a notable home market effect occurs.

As mentioned above, factor mobility crucially enters NEG theory. In fact, both the location of production, which underpins capital mobility, as well as labour migration flows are essentially central to the NEG story. Although empirical research using a NEG framework is not particularly
rich on these topics, some authors attempt to test theoretical predictions in these areas. Head and Mayer (2004a) study the location behaviour of Japanese firms investing in the EU adopting a model of location choices coherent with the theoretical setting of NEG. It emerges clearly that demand matters for the location of production confirming that market access considerations represent a strong driver of capital mobility. Similarly, Crozet et al. (2004) find evidence of demand linkages shaping FDI in France while LaFountain (2005) confirms the importance of proximity to large markets in firms’ location choices in the US just for some sectors of manufacturing activity. Okubo et al. (2010) consider the effect of firms’ heterogeneity on location choices in a setting of market integration. This study suggests that for decreasing trade costs more efficient firms tend to agglomerate in the core, where demand is higher, while less efficient firms cluster in the periphery because of the less intense competition in this location. However, in Okubo et al. (2010) this relationship between economic integration and localisation patterns appears not to be monotonic, as suggested instead in Baldwin and Okubo (2006). In fact, as integration deepens further, market access becomes the main pull for firms’ location choices, as suggested by above-mentioned studies. Indeed, the effect of protection from competition arising from locating in the periphery gradually loses relevance as trade costs become lower and lower. As a consequence, less efficient firms modify their location behaviour by setting up in the core, where market access is relatively higher. With respect to migration flows, Crozet (2004) tests the occurrence of forward linkages (i.e. workers are attracted by location with large production) in determining agglomeration in Europe. This contribution suggests that forward linkages do effectively matter but, in the case of Europe, low labour mobility fundamentally impedes the exacerbation of the core-periphery pattern. Similarly, Pons et al. (2007) confirm the relevance of forward linkages for the attraction of migration flows and the consequences on the spatial distribution of economic activity in Spain, whereas d’Artis Kancs (2011) structures a model suggesting that the access of Eastern countries to the EU determines a net (but low) migration of workers from East to the West, as predicted by NEG. In general, it emerges that NEG theoretical predictions have become to be tested in empirical studies in recent years and this represents a further step towards a better understanding of the implications of economic geography on the spatial structure of economic processes.

7.2 NEG and EU enlargements to neighbouring countries

Other contributions adopting NEG-related concepts explore the importance of various drivers of industry localisation, with particular reference to the case of closer European integration (Midelfart-Knarvik et al., 2000; Forslid et al., 2002a, 2002b; Midelfart-Knarvik and Overman, 2002; Marques, 2005). What most of these studies generally suggest is that deeper integration has specific consequences on the spatial distribution of economic activity. In fact, agglomeration forces seem to drive the localisation of European industry towards few locations as integration becomes tighter. Moreover, sectoral differences arise with capital-intensive and skill-intensive activities concentrating
in the core of the EU (sometimes after an initial dispersion) while slow growing industries characterised by unskilled labour tend to agglomerate in peripheral areas. A similar story is told also by Brülhart et al. (2004) and Crozet and Koenig (2004), who study the market access effect of European integration on the location behaviour of economic activity. These studies suggest that European regions close to the border with new member countries will benefit by attracting industry due to market access related advantages and cheaper imports. In general, what is suggested by most is that tighter economic integration in Europe could plausibly trigger agglomeration processes that lead to divergence of income across regions (see Marques, 2008). These kinds of dynamics of divergence and polarisation are also outlined by Petrakos (1996; 2000) and Bradley et al. (2005) with respect to transition economies in a framework which is not strictly NEG-related. As a matter of fact, such studies argue that the rapid internationalisation of the economy of CEECs and the following integration in the European single market basically results in a disproportionate agglomeration of economic activity in metropolitan regions of CEECs (Petrakos and Economou, 2002) as well as in regions bordering the EU due to better market access. As a consequence, economic polarisation and divergence appears to be associated with closer economic integration in the case of EU enlargement eastward. This view is supported by a series of empirical studies in Traistaru et al. (2003), who find evidence that the process of economic integration of CEECs with the EU has translated into within-countries relocation of industry to the benefit of CEECs capital regions as well as areas bordering the EU, where agglomeration economies and market access considerations, respectively, dominate. Therefore, growth prospects for these winning regions seem relevant while other losing regions are expected to stagnate or decline. In the same debate, Damijan and Kostevc (2011) provide evidence that in most CEECs there exist a U-shaped relationship between economic integration and regional divergence. In other words, initially increasing trade liberalisation sharpens inequalities in relative wages due to strong agglomeration effects. In this phase, developed regions in CEECs disproportionally benefit from economic integration as compared to less favoured regions as a result of strong agglomeration economies which attract industry. As such, these regions exhibit higher wages relative to other regions and polarisation occurs. In a second step, however, Damijan and Kostevc (2011) suggest that in most (but not all) CEECs the growing inflow of foreign capital in regions bordering the EU lead to a fast process of adjustment of regional wages which foster convergence. The debate about regional growth and convergence-divergence patterns in CEECs within the framework of economic integration is further enriched by Monastiriotis (2011) who suggests that such dynamics are particularly complex. Indeed, he highlights that regional growth is far from being a linear process and that neoclassical convergence, cumulative causation leading to divergence, and non-monotonic convergence as a function of national development à la Kuznets may coexist. Overall, a divergence path of economic development seems to dominate in CEECs leading gradually to a pattern of polarisation (Monastiriotis, 2011). Moreover, economic integration has implications also in terms of regional specialisation. In the case of closer integration between the
EU area and transition economies this adds a dimension to the process of economic restructuring of industry in these countries. Different country-level and comparative studies in Traistaru et al. (2003) note that regional specialisation increases in some CEECs while decreases in others. In some countries important interregional shifts in industry location are found over time (see for example the chapter by Damijan and Kostevc in Traistaru et al., 2003) and they underpin the concentration and specialisation of industry in some regions while other areas become more diversified. In a study on the effect of integration on the structure of industry in CEECs, Kancs (2007) suggests that specialisation has decreased on average. The apparently ambiguous effect of integration on regional specialisation, however, seems to be justified by the planned economy inheritance of CEECs, according to which regions were in most cases specialised in activities uncorrelated with local comparative advantages. Such a distorted pattern of specialisation might now be under restructuring and CEECs regions may presumably experience a re-specialisation due to transition and integration with the EU. Hence, the new pattern of re-specialisation could justify the fact that not only some CEECs exhibit increasing regional specialisation whilst others a decreasing trend, but also that average industrial specialisation is reduced by closer economic integration with the EU.

In general, most contributions suggest that European economic integration with transition economies determines a set of changes in the economic geography of the latter leading to sectoral restructuring at both intra- and inter-country level (Traistaru et al., 2003) as a result of variations in the drivers of firms’ location behaviour (Baldwin and Wyplosz, 2006). Eventually, as outlined above, such a restructuring is frequently deemed to determine divergence patterns between CEECs and old EU members as well as regional polarisation within CEECs (Petракos et al., 2005; Krieger-Boden and Traistaru-Siedschlag, 2008).

8. Concluding remarks

This paper summarizes the main insights offered by NEG with respect to economic integration between countries and regions. What emerges from reviewing the theoretical framework of NEG is a fundamental ambiguity in the response of spatial economic processes to the gradual removal of trade barriers. In fact, as highlighted in the text, most (but not all) NEG models predict a bell-shaped association between the agglomeration of economic production and welfare in a few places and the intensity of trade liberalisation. As such, with low economic integration the spatial distribution of industry and income is rather dispersed, with each market served locally. However, the gradual removal of trade constraints triggers self-reinforcing agglomeration processes: as a result, industry tends to concentrate in few places. Finally, when economic integration is extremely strong, dispersion forces prevail and factor and product market competition basically drive the geographical distribution of production. Attaining such a deep degree of economic integration, however, is not a simple task. As a matter of fact, barriers to trade are not only represented by ‘natural’ trade obstacles
such as tariffs and quotas, but also by other elements such as different regulatory frameworks as well as different languages and cultures (Krugman and Venables, 1990). Therefore, full economic integration appears to include a wider number of non-economic elements. As we have discussed, most NEG is concerned with the effects of heterogeneous locations on the decisions of a uniform mass of economic actors. In this respect, it is worth highlighting that recent directions in research point out that ‘micro-heterogeneity’ across workers and firms is likely to affect the occurrence, the strength and the distributional effects of agglomeration economies (see for instance Okubo et al., 2010 and Fallah et al., 2011), thus encouraging the analysis of the effects of the interaction between actors’ and locations’ heterogeneous characteristics on centripetal and centrifugal forces (Ottaviano, 2011). With respect to empirical research, we argued that such aspect of NEG is not yet fully developed and we mentioned some relevant contributions aimed at investigating the occurrence and the importance of NEG forces, ranging from the home market effect and market access considerations to the dynamics of capital and labour mobility. We also reviewed relevant empirical studies concerned with the effect of the EU enlargement eastward. Most of these works seem to suggest that economic integration lead to a restructuring of industry in CEECs and that relocation patterns characterise most of the economic geography of such countries. Divergence and polarisation between regions in new EU member countries appear to be among the main consequences of enlargement, with more favoured regions (metropolitan and regions bordering EU) taking off while others stagnating or declining. But how does NEG enter the domain of policy? Clearly, NEG theoretical predictions tend to be rather catastrophic so that policy implications are problematic to draw. Think for instance to cumulative agglomeration processes whose result is the concentration of production in just one location. Nevertheless, this is mainly due to the high level of abstraction of theoretical models which tell nothing about real geography, the role of different institutional settings and history, among others. Generally, as explained by Ottaviano (2002) one major policy implications of NEG is the understanding that all sort of policies and interventions most likely have regional side effects influencing agglomeration and dispersion forces. As such, a careful evaluation of such measures should also include an analysis of the potential effects which might reinforce or alleviate centripetal and centrifugal forces. In other words, it appears crucial to understand all the factors and dynamics that can play a role in the process of integration, ranging from institutional settings to initial conditions and local economic geography (Bradley et al., 2005), in order to define policies that are able to tackle the negative effects of the agglomeration process through “spatially targeted incentives in specifically designated areas” (Traistaru et al., 2003).
9. References


Ohlin B., 1933. *Interregional and international trade*. Harvard University Press.


