10/03/05:

New Economic Geography*

Anthony J. Venables,
London School of Economics and CEPR

Abstract:
New economic geography provides an integrated and micro-founded approach to spatial economics. It emphasises the role of clustering forces in generating an uneven distribution of economic activity and income across space. The approach has been applied to the economics of cities, the emergence of regional disparities, and the origins of international inequalities.

* Written for Palgrave Dictionary of Economics

Author's address:
A.J. Venables
Dept of Economics
LSE
Houghton Street
London WC2A 2AE, UK
a.j.venables@lse.ac.uk
http://econ.lse.ac.uk/staff/ajv

1. Introduction

Why is economic activity distributed unevenly across space, with centres of concentrated activity surrounded by 'peripheral' regions of lower density? What economic interactions are there between different geographical areas, and how do these shape income levels in the areas? How does the spatial organization of economic activity respond to exogenous shocks, such as technological change or policy measures? The contribution of 'new economic geography' (NEG) is to address these questions in a manner that is based on rigorous micro-economic foundations. It shows how the spatial structure of an economy is determined by the interplay between costs of transactions across space and various types of increasing returns to scale. The questions posed above can be addressed at different spatial levels, international, regional and urban. NEG provides a unified framework for analysis at these different levels.

2. Clustering vs dispersion.

There are several key analytical ingredients of the NEG approach. The first is the recognition that spatial interactions are costly. These costs are shaped by geography and depend on the nature of the interaction. Thus, trade in goods incurs shipping costs and costs of time in transit, these depending on distance shipped, on transport infrastructure and on geography. Communications and coordination costs mean that workers may be less effective if they are not in close proximity with co-workers. Factor mobility may be impeded by distance and geography. This approach contrasts with that of international trade theory, in which spatial units are identified solely with countries – jurisdictions rather than geography – and where goods and factors are typically assumed to either be traded freely, or to be completely non-tradable. The NEG approach shows how outcomes depend on the extent to which different goods and activities are mobile between locations.

The second key ingredient is the possibility that there are clustering forces, inducing activity to concentrate in space. Clustering arises because of spatially concentrated increasing

returns to scale which can derive from a number of different underlying forces.[1] One possibility is that there are public goods, the enjoyment of which depends on geographical access, such as a town centre. Another possibility is that there are positive technological externalities such as knowledge spillovers; firms produce ideas that can be observed and copied by other firms, depending on their proximity. These approaches have been prominent in much of the urban economics literature (eg Henderson 1988), but writers in the NEG literature have generally sought to derive clustering forces from spatial interactions in imperfect markets, rather than to simply assume them through public goods or technological externalities.

One way to derive clustering forces is through thick market effects, particularly in the labor market. Dense labor markets may allow for better matching of the skills of workers and the requirements of firms (Helsley and Strange 1990). Incentives to acquire skills may be greater where workers face more prospective employers (Matouschek and Robert-Nicoud 2005). Another way in which to derive clustering is to use industrial organization models of imperfect competition. The route followed in much of the NEG literature is to suppose that an industry (we will call it 'manufacturing') contains a number of firms, each of which has increasing returns to scale. The presence of internal economies of scale means that firms are faced with a location choice (if they had constant or diminishing returns then, given transport costs and dispersed consumers, they would choose to produce a very small amount in all locations – 'backyard capitalism' Starrett 1978). The questions are then, where do firms choose to locate, and under what circumstances will they cluster together? The model often used to analyse the choice is the Dixit-Stiglitz (1977) model of monopolistic competition and its international trade extensions (Krugman 1980). In this model each firm has a distinct variety of product which it produces in a single location and exports to other locations, and entry and exit occur until profits are bid down to zero. It turns out that as firms take location decisions in order to maximize profits, so their location pattern tends to amplify any underlying differences between locations, and from this it is possible to generate an outcome in which clustering occurs.

To understand the argument, suppose that there are two regions A and B, and that A has demand $k > 1$ times larger than B (ignoring factor supply considerations for the moment). Could

[1] The classic discussion is Marshall (1890), and for a recent survey see Duranton and Puga (2004)

there be an equilibrium in which firms are located in proportion to the size of the regions, so A has *k* times more manufacturing firms than B? If trade costs are prohibitively high the answer is yes; only local firms supply each market, and the number of firms is proportional to the size of the market.[2] But as trade costs are reduced and firms start to export, two things happen. First, the region B market comes to be supplied by *k* times as many importing firms as does the country A market, this reducing the profitability of producers in B. Second, each firm in B will pay transport costs on a large part of their output (sales to the large country A market) while firms in A will only pay transport costs on a smaller fraction of their output (sales to the smaller region B market). Both arguments suggest that firms in A become relatively more profitable, implying that in equilibrium with free entry the number of firms in A must exceed the number in B by a factor greater than *k*. The large region therefore has a disproportionately large share of manufacturing production, and is a net exporter of manufactures and importer of agriculture. More generally, a region with good 'market access' will attract a high share of firms.

This argument holds only if transport costs lie strictly between zero and a prohibitive level. If transport costs are prohibitive no firms ship any exports; each region is self sufficient, and the location of industry is in proportion to the size of the regions. Conversely, if transport costs are zero, then the argument collapses, as firms in all regions have equally good access to all markets. The argument shows that it is at intermediate levels of transport costs that market access matters, and manufacturing is pulled disproportionately into the large region.

While this argument creates an incentive for clustering of firms, it is balanced by dispersion forces. These could be due to negative externalities, such as congestion, or arise as a consequence of immobility of some factors of production. Which factors are immobile depends on context, but typically include land (as in the tradition of urban economic modeling) and some or all types of labour. Thus, if labour were immobile, any benefit that firms derive from locating in one region rather another would create a regional wage differential, until profits (more generally, the return to mobile activities) are equalized across regions.

Labour mobility is central to the Krugman (1991) 'core-periphery' model. This analyses two regions and two sectors, a constant returns to scale agriculture and manufacturing modeled

2 Notice that this argument uses the Dixit-Stiglitz property that all firms are the same size in equilibrium.

as was outlined above. Each sector uses a sector specific type of labour ('peasants' and manufacturing workers respectively), and the regions' endowments of these factors are, ex ante, identical. Crucially, manufacturing workers are mobile between the locations, whereas peasants are immobile. What is the division of manufacturing workers and firms between the two locations? Outcomes, as a function of trade costs, are illustrated on figure 1. When trade costs are high manufacturing is equally divided between regions. However, when trade costs are low enough, manufacturing (and all manufacturing workers) concentrate entirely in one region or the other. There are two mutually reinforcing arguments supporting this clustering. The concentration of manufacturing workers creates a large market, so making the location profitable for firms. And the entry of firms bids up wages, so making the location attractive for workers (this effect reinforced by the fact that workers also benefit from not having to pay trade costs on their consumption of manufactures). It is not profitable for any single firm to leave the cluster, because the benefit of lower wages is outweighed by the loss of market access. As the figure makes clear, the switch from dispersed manufacturing to agglomeration arises discontinuously. There is a critical value of trade costs, t^*, above which dispersed production is the stable equilibrium, and below which dispersed activity is unstable, while clustering of activity, in either one of the regions, is a stable equilibrium.

Krugman's 'core-periphery' model is perhaps the seminal paper, and brings the insight that agglomeration forces can be derived from a standard model of trade and monopolistic competition.[3] These micro-foundations mean that outcomes (clustering or dispersion) can be linked to parameters such as trade costs, as in figure 1. The model also makes it clear that ex ante identical locations can be different ex post, and that there are multiple equilibria – we have to look outside the model, or rely on chance, to determine which of the regions has the manufacturing cluster.

The model was constructed with just two locations. How do these insights extend when there are many locations? With many locations the number of equilibria increases dramatically, and there is a danger that little can be said about outcomes. There are several ways through this problem. One is to investigate how the size and number of manufacturing centres on a given

3 See Fujita, Krugman and Venables (1999) for development of many of these models.

geographical space depends on underlying parameters such as trade costs and population levels. The approach of Fujita, Krugman and Venables (1999) is to hypothesise a circular economy (with population on the circumference) and show that an initial random allocation of manufacturing grows into a determinate number of centres, the size of which are greater (and number of which smaller) the lower are trade costs. Given some number of centres, reducing trade costs will have no effect until some critical point is reached, at which the economy will reorganize itself to a new economic geography with fewer and larger centres. The approach of Fujita and Mori (1997) is to suppose that initially there is a small populated region. Population growth causes this to expand, at first with the spread of agricultural production into the hinterland. However, these agriculture workers demand manufactures, and this will cause new manufacturing centres will develop. The expanding economy therefore grows its urban structure, and cities will tend to be larger (and further apart) the greater are increasing returns to scale and the lower are trade costs. Both of these approaches work with underlying geographies that are undifferentiated. Adding structure to these underlying geographies simplifies the problem in fairly natural ways. A transport node – such as a port or river crossing – will attract manufacturing, as firms in such a location have better access to a larger number of consumers.

3. Intermediate goods and industrial clusters

The clustering mechanisms described in the preceding section turn on the mobility of labour. Clustering occurs because as firms and workers move, so do both supply *and* demand for manufactures. What if labour is immobile? An analogous mechanism can work between firms when we take into account intermediate goods, i.e. goods that are both supplied and demanded by the manufacturing sector. This mechanism is similar to the idea of 'linkages' common in the development economics literature of the 1950s and 1960s. This studied the roles of backward linkages (demands from downstream firms to their suppliers) and of forward linkages (supply from intermediate producers to downstream activities) in developing industrial activity. However, as we saw above, rigorous treatment requires that the concepts are placed in an environment with increasing returns to scale, in order to force firms to make a location choice.

This can be done in a model isomorphic to that outlined above, but in which firms in the manufacturing sector produce and use intermediate as well as final goods. Clustering can occur as it is profitable for firms producing intermediate and final goods to co-locate. Depending on the strength of linkages within and between industrial sectors, clustering might occur through a wide part of the economy, or within narrowly defined sectors.

In this model clustering arises purely from the mobility of firms, even if there is little or no labour mobility. It is applicable to a number of different situations. For example, within a country there might be inelastic supply of land or housing in each city which places a limit on labour mobility. Clustering of particular sectors can nevertheless occur, and might be associated with different levels of employment and different house prices across cities.

The model has also been applied in the international context, with labour immobile across national boundaries. Manufacturing may then concentrate in a single country or group of countries, and this clustering may lead to international wage differences. This idea is developed by Krugman and Venables (1995) in a model with two countries, N and S, assumed to be ex ante identical. Firms produce final and intermediate goods, and use labour and intermediates as inputs. Equilibrium outcomes are summarized in figure 2, which has trade costs on the horizontal axis and real wages on the vertical axis. At very high trade costs there is no clustering, so the two economies are identical; this is because firms operate in each country to supply local consumers. As trade costs fall (moving left on the figure) so the possibility of supplying consumers through trade rather than local production develops, and clustering forces become relatively more important. Below some level of trade costs, t^*, clustering forces come to dominate, and one of the countries (N in the figure) gains most of manufacturing, and consequently has a high real wage. This clustering 'deindustrialises' the other country (S) which experiences a fall in its real wage. For the case illustrated in figure 2, there is a range of trade costs in which the world necessarily has a dichotomous structure. Wages are lower in S than in N, but it does not pay any firm to move to S as to do so would be to forego the clustering benefits of large markets and proximity to suppliers that are found in N. However, as trade costs fall it becomes cheaper to ship intermediate goods, so the location of manufacturing becomes more sensitive to factor price differences. This is the era of globalisation, in which

manufacturing starts to move to S and the equilibrium wage gap narrows. In this model factor price equalization is attained when trade is perfectly free -- the 'death of distance'.

This model offers quite a general theory of location, in which four forces are at work, two of which are dispersion forces, and two favour clustering. The dispersion forces are *factor supply* and *product market competition*: moving a firm from S to N reduces the profitability of firms in N both by bidding up wages and by driving down product prices. Against this there are two agglomeration forces, *demand linkages* and *cost linkages*: moving a firm from S to N raises the profitability of firms in N by increasing the size of the market and by increasing the supply of intermediate goods. The balance between these four forces depends on parameters, including trade costs, giving the outcomes illustrated on figure 2. It is worth comparing the four forces present in this model with the conventional model of free international trade, in which factor supply alone determines the location of economic activities.

Extensions of this approach provide a number of further insights concerning international inequalities. It suggests that the world may tend to organise into a rich club of countries and a poor club. Economic development takes the form of countries growing from the poor club to the rich club in sequence, rather than in parallel. Parallel growth is unstable, because of the tendency of developing manufacturing sectors to cluster in a few countries.

4. Empirical findings.

The new economic geography literature offers explanations of a number of phenomena that are empirically well documented – even obvious – such as the existence of cities, and the presence of regional and international inequalities. Its insights range across different spatial scales, from the urban to the international. Empirical work is correspondingly diverse, and we refer to just four elements of it.

First, there is strong evidence of the importance of geography in shaping economic interactions. Trade costs are high (Anderson and van Wincoop 2004), and 'gravity modelling' points to the fact that bilateral trade flows approximately halve with each doubling of distance between country pairs. Similar results hold for other cross-border interactions such as foreign

direct investment flows, telephone calls, and international portfolio investments.

Turning to outcomes, a number of researchers have investigated the extent to which individual sectors are prone to clustering. There is a long business school tradition of work in this area, for example Porter (1990), who studies a number of industrial clusters. Econometric work has established that sectors are more prone to cluster than would be explained by chance or by comparative advantage (Ellison and Glaeser 1997). A further prediction of NEG is that prices of immobile factors will be high in locations with good market access. As we have seen, in the national context this will show up in the price of land and housing and hence nominal wages differences, a prediction confirmed for US counties by Hanson (2005). In the international context this may show up as real wage differences. Gallup and Sachs (1999) find that 70% of cross-country variation in per capita income can be accounted for by just four measures of physical and economic geography (malaria, hydro-carbon endowment, coastal access and transport costs). A structural approach to identifying the importance of market access in explaining cross-country income differentials is adopted by Redding and Venables (2004) who use gravity modeling to calculate measures of market access for each country. Controlling for other factors (such as institutional quality), these measures of market access are important determinants of international wage gaps.

Finally, there is considerable evidence of the productivity benefits derived from being located in dense centres of economic activity. A recent survey of the literature on cities (Rosenthal and Strange 2004) reports a consensus view that doubling city size is associated with a productivity increase of some 3 – 8%. However, a good deal of uncertainty surrounds the extent to which this is driven by the different clustering mechanisms – knowledge spillovers, thick labour markets, market access benefits, or inter-firm linkages -- that we described above. Identifying the importance of each of these underlying mechanisms remains an active area of current research.

References

Anderson, J. and E. van Wincoop (2004) 'Trade costs', *Journal of Economic Literature*, 42, 691-751.

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

Duranton G. and D. Puga (2004) 'Micro-foundations of urban agglomeration economies' in V. Henderson and J. Thisse (eds) *Handbook of Urban and Regional Economics*, vol 4, North Holland, Amsterdam.

Ellison, G. and E. Glaeser (1999) 'Geographic concentration in US manufacturing industries; a dartboard approach', *Journal of Political Economy*, 105, 889-927.

Fujita, M., P. R. Krugman and A J. Venables (1999), *The Spatial Economy: Cities, Regions and International Trade*. Cambridg.e, Mass.: MIT Press.

Fujita, M., and T. Mori (1997) 'Structural stability and the evolution of urban systems' *Regional Science and Urban Economics*, 27, 399-442.

Gallup, J.L and J. Sachs (1999) 'Geography and economic development' in B. Pleskovic and J.E. Stiglitz (eds) '*Annual World Bank Conference on Development Economics; 1998*', World Bank, Washington DC.

Hanson, G. (2005) 'Market potential, increasing returns and geographic concentration', *Journal of International Economics*, 67, 1-24.

Helsley, R.W. and W.C. Strange, (1990) 'Matching and agglomeration economies in a system of cities' *Regional Science and Urban Economics*, 20, 189-212.

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

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

Krugman, P. R. (1991), 'Increasing Returns and Economic Geography' *Journal of Political Economy*, 49, 137-150.

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

Marshall, A. (1890), *Principles of Economics*. London: Macmillan (8th edition, 1920).

Matouschek, N. and F. Robert-Nicoud, (2005) 'The role of human capital investments in the location decisions of firms', *Regional Science and Urban Economics*, 35 (5), 570-583

Porter, M.E., (1990), *The competitive advantage of nations*, New York, Macmillan.

Redding, S, and A.J. Venables (2004) "Economic geography and international inequality", *Journal of International Economics*, 62, 53-82.

Rosenthal, S.S. and W.C. Strange (2004) 'Evidence on the nature and sources of agglomeration economies', in V. Henderson and J. Thisse (eds) *Handbook of Urban and Regional Economics*, vol 4, North Holland, Amsterdam.

Starrett, D.A. (1978) 'Market allocations of location choice in a model with free mobility', *Journal of Economic Theory*, 17, (1), 21-37.

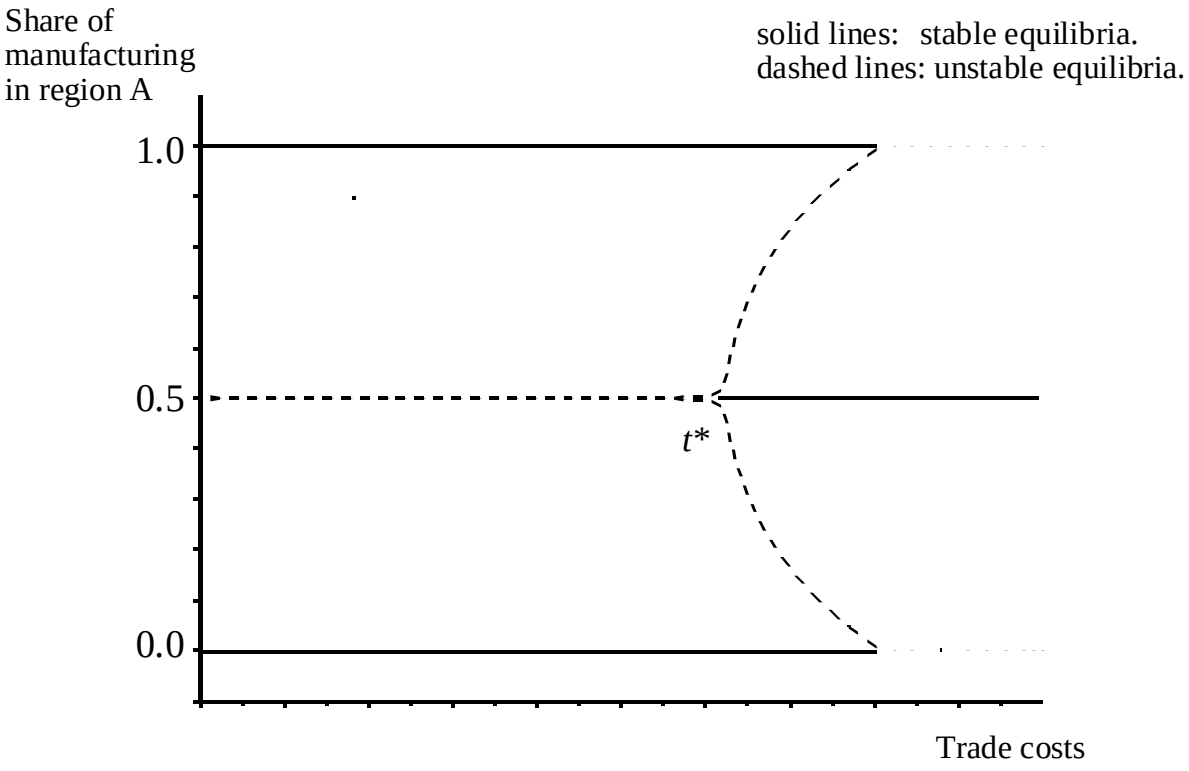

Fig 1: Location of manufacturing in two regions

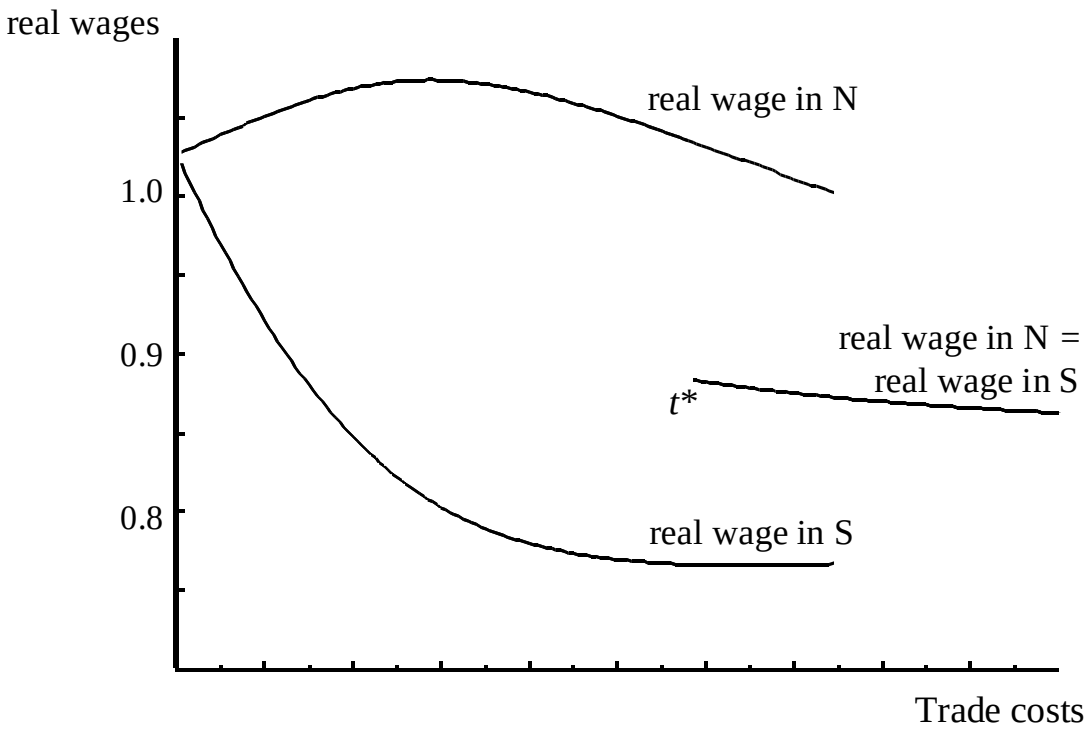

Fig 2: Real wages in a two country model