SWEDISH ECONOMIC POLICY REVIEW 12 (2005) 143-169

European integration: A view from geographical economics

Anthony J. Venables*

Summary

■ This paper reviews aspects of European integration through the lens of new economic geography. It analyses factors shaping the location of industry in the European Union, and studies the evolution of spatial differences in income levels. Recent years have seen a narrowing of income differentials between countries and a widening of differentials between regions within countries. Much of this is due to the resurgent prosperity of cities in the EU. The paper concludes with speculation about possible future development of the European city system. ■

JEL classification: F1, F15.
Keywords: European integration, economic geography.

** Anthony J. Venables is Professor at London School of Economics and a Research Fellow at CEPR.*

SWEDISH ECONOMIC POLICY REVIEW 12 (2005) 143-169

European integration: A view from geographical economics

Anthony J. Venables*

Economic evaluations of European integration have posed two main questions. What are the economic gains—or losses—from integration? And how are they distributed between member countries and regions? Answering these questions requires that researchers identify the channels through which integration affects the economy. One of the most important of these is that integration changes the attractiveness of different locations for different production activities. In response to this there are changes in labour demands and real incomes between regions. If labour is mobile then migration flows may follow, changing the distribution of population between regions and cities. Economic geography, with its cumulative causation mechanisms, suggests that these changes may potentially be quite large. The objective of this paper is to outline some of the possibilities, and review the experience of the EU in the light of these possibilities.

Economic analysis of these issues has gone through many generations of steadily improving models. Advances have been driven largely by innovations of method in the academic literature, but also by the realisation that analysis of the effects of deep integration amongst a group of similar countries requires tools different from those of traditional trade theory. The first generation approach focussed on trade creation and trade diversion, raising doubts even about the presence of aggregate gains (Viner, 1950). The point is that regional integration agreements (RIAs) cause countries to specialise according to their regional comparative advantage, not their global one. Consequently a country that is labour scarce relative to the world but labour abundant relative to its RIA partners may find its production structure shifting in the "wrong " way as it expands trade within the RIA. However, the real income damage that this causes is typically felt not by this country, but by other countries in the RIA. Trading part-

* *Thanks to Martin Stewart and Holger Breinlich for research assistance and Karolina Ekholm for valuable comments.*

ners in the RIA have their imports diverted in line with intra-RIA comparative advantage, rather than worldwide comparative advantage (Venables, 2003).

The new trade theory of the last quarter of the 20th century focussed on firms, on increasing returns to scale, and on imperfect competition. This brought a more optimistic picture of the gains from regional integration. Internal trade liberalisation intensifies competition. Firms respond by merging or going bust, leaving a resultant industrial structure in which average costs are lower (as remaining firms are larger and better exploiting economies of scale) and prices are closer to marginal cost (as, even after exit or merger, the resultant market structure is more competitive). An additional mechanism touched upon in this literature (Flam, 1992) is now the subject of a tidal wave of current research following from the work of Melitz (2003). Firms in an industry typically have widely differing levels of technical efficiency, and those that are forced out in this process of rationalisation are generally the least efficient. This amplifies the reduction in average costs associated with integration, so increasing the aggregate gains.

Increased focus on firms also brought greater attention to bear on their location decisions, and hence on the economic geography of an integrating region. If integration causes a rationalisation of industrial structure, which regions are likely to gain firms, and which are likely to loose them? The first observation was that locations with good market access would tend to be relatively attractive for firms, and that this would bid up wages in these locations. The second was that ensuing high wages would attract labour inflow, increasing market size and reinforcing the market access advantage of the region. From this positive feedback the core-periphery model was born (Krugman, 1991a). The third observation, was that market access derives not only from consumers' final demands but also from firms' intermediate demands, so firms producing intermediate goods tend to locate close to their downstream customers, and the downstream customers in turn want to locate close to intermediate suppliers (Venables, 1996). Clusters of industrial activity therefore form. Other agglomeration mechanisms—dating back to Marshall (1890) and much studied by economic geographers—reinforce this process of clustering. In all these models integration is likely to promote the spatial agglomeration of activities.

This brief review indicates how the literature now gives a rich set of tools to address the effects of economic integration on industrial location in particular, and on other economic variables more widely.

What do we make of the European experience in light of this literature?

1. Industrial location

1.1. Long run outcomes

Given a map of Europe with endowments in place (natural resources, labour, and infrastructure) but no pre-existing industrial structure and no pre-EU trade barriers, where would various economic activities locate? It is easy to write down a list of considerations that are of varying degrees of importance to firms in different sectors.

At the top of the list is market access. For the vast majority of activities in the modern economy access to consumers is the overwhelmingly important factor, and these activities go more or less proportionately to where the people are. Such sectors are of course personal services, government, health, retailing, etc; these activities simply locate where their market is. Market access matters more or less to other activities, according to what we may somewhat loosely call their "transport intensity". Of course, low cost access to markets is important for all firms, but it is relatively more important for sectors with higher transport intensity, and hence higher costs of shipping goods to final consumers.

Next on the list is access to inputs of production, the stuff of Heckscher-Ohlin trade theory. Low cost inputs are valuable to all producers, but costs of some inputs matter more to some sectors than others, depending on the sector's "factor intensity". The simpler cases of this are the basis of much textbook trade theory, and we know exactly how factor endowments shape industrial location even though—at equilibrium—there may be no spatial variation in factor prices.

Third on the list is the profitability of locating close to complementary activities. This applies within the firm, as there are likely to be "disintegration costs" associated with spatially fragmenting activity. It also applies between firms, driven by a number of mechanisms. One is the dual relationship of cost and demand linkages; downstream firms' proximity to firms supplying intermediate goods, and upstream firms' proximity to firms that use their output. Another is the desirability of locating close to firms that are innovating in the hope of benefiting from technological spillovers. A third is the fact that firms in a sector may train workers in specialist skills that are then valuable to neighbouring firms.

Given these factors, would we be able to complete the map, identifying what activities are performed where? The answer is yes, up to a point. There are two qualifications. The first is that the assignment of industries to locations is not in general unique, for two distinct reasons. One is that, as is well known to students of higher dimensional trade theory, if there are more industries than factors of production, then the location of industry is indeterminate. Essentially, full employment of all factors in a particular location can be achieved by many different combinations of activities. This non-uniqueness can apply at the level of quite broad sectors, and much more so for narrowly defined activities. The other reason is that if firms' location is determined by proximity to complementary activities then clustering may occur. Firms will locate in the cluster in order to benefit from complementarities but—within limits—it does not matter exactly where the cluster is located. No firm wants to move out of the cluster given that other firms are there, but that could be true regardless of the precise location of the cluster.

The second qualification is that, even if location of activities is in principle unique, it might depend on the underlying forces of market access and factor endowments in a complex way. The key insight to understanding the dependence is simply comparative advantage, but operationalising this concept can be very much more complex than the textbooks suggest. The problem is that the full list of industry characteristics—skilled labour intensity, capital intensity, R&D intensity, intermediate input intensity, transport intensity etc.—is very long, as is the associated list of country characteristics. To describe the actual equilibrium location of industry the simultaneous interaction of all these forces needs to be taken into account.

Furthermore, any particular firm or plant is a composite of different activities with different characteristics, each of which might be best located in a different place. Generally activities come packaged up—"disintegration costs" bind these activities together in one location; a plant contains both skilled labour intensive and unskilled labour intensive production processes, and decisions have to be based on the combined whole. These "disintegration costs" may be the need for continuous production flow, transport costs, or more complex considerations such as management or coordination problems in splitting production. Yet sometimes "disintegration costs" may be low enough for an activity to be able to fragment into component elements, all of which may then relocate in response to comparative advantage. The analytical difficulty arises because whether fragmentation takes place or not is an endogenous decision, depending on the level of disintegration costs relative to the production cost saving achieved by fragmentation.

In a world where there are several sources of comparative advantage and the possibility that productive activities can fragment in this way, what can be said about the pattern of specialization and trade? This question is explored by Markusen and Venables (2004). The authors suppose that just two elements of comparative advantage are in operation. One is the interaction between the factor abundance of countries and the factor intensity of goods, and the other is the interaction between the "remoteness" of countries and the transport intensity of goods.[1] Since countries are described by their factor abundance and remoteness they can be placed in the two dimensional space of Figure 1. The horizontal axis of the figure measures factor abundance (let us call it the capital/labour ratio), while the vertical is remoteness, as measured by countries' transport costs. This is a measure of the transport costs each country faces when trading with all others (it is expressed as a proportion of production costs, and assumed to be the same for all goods—whether imported or exported—and for trade with all of a country's trading partners). Thus, point *E* on the figure represents a particular country; as illustrated, country *E* is capital abundant (towards the left of the figure) and quite remote (towards the top of the figure, with high trade costs). The example is constructed with a country at every point on a fine grid across the surface of the figure.

[1] See Venables and Limao (2002) for analytics of this in a case without fragmentation.

Figure 1a. Production and trade

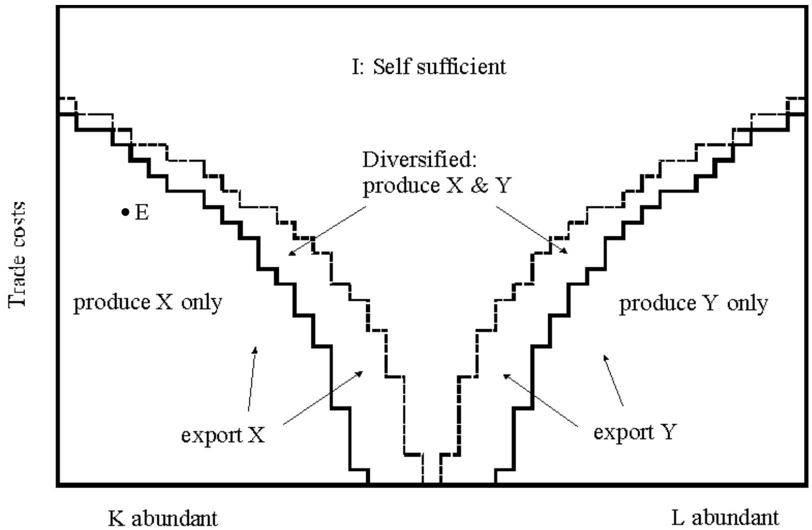
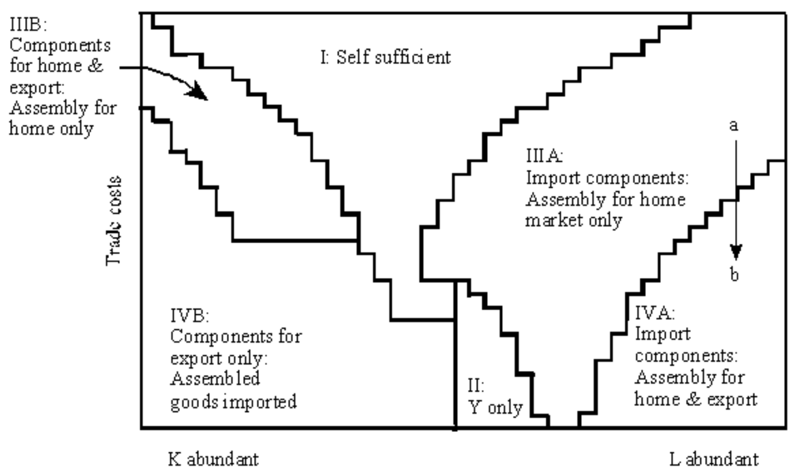


Figure 1b. Component production, assembly and trade



Initially each of these countries can engage in two activities; production of a relatively capital intensive good, *X*, and a relatively labour intensive good, *Y*. Supposing that the two goods are equally transport intensive, then equilibrium production patterns of countries at each point are as given in Figure 1a. Countries' production structures are determined by their factor endowments and remoteness, with specialisation being more likely the more extreme the factor endowment and the less remote the country. The example is constructed to be symmetric, but the effect of changing this assumption can easily be seen. If, for example, good *Y* were less transport intensive than good *X* then the boundaries on the right hand side of the figure would become steeper.

What happens if we extend this analysis to the case where it becomes possible to "fragment" production of one of the activities? Let us suppose that *X* production can be split into assembly (*A*, which is the most labour intensive of all activities) and component production, (*C*, the most capital intensive). Furthermore, let us suppose that there are "disintegration" costs, these arising as shipping components is costly. Thus, if *X* production is fragmented it is able to relocate its elements to minimise production costs, but at the same time it becomes relatively transport intensive, as transport costs are incurred both in shipping component parts and in shipping final assembled output.

In the full equilibrium, who does what? The different areas on Figure 1b illustrate outcomes. Countries in region I do not trade at all (they have no inter-industry trade and in this simple model no intra-industry trade takes place). They are self-sufficient because they have relatively high trade costs and their factor endowments are close to the factor intensities of *Y* production and of combined *X* production (components and assembly). Countries in region II specialise in production of good *Y*, as their factor endowment ratio is very close to the factor intensity of *Y*.

Regions III and IV are the main focus of our analysis. In region IIIA countries are labour abundant but have quite high trade costs. Labour abundance means it is profitable for them to undertake assembly of *X* using imported components. But trade costs make it unprofitable for them to export the final product; this would be highly transport intensive, as components would bear trade costs twice—on import and on export embodied in the final product. Components are therefore imported for assembly just for domestic sales. This case corresponds to "horizontal" or "market serving" international production.[2] Region IIIB is the capital abundant counterpart to this. Capital intensive components are exported; but countries in this region have high enough trade costs that it is profitable for them to undertake assembly, at least for the local market.

Regions IV are areas of the world where full "fragmentation" of production takes place. In region IVA labour abundant countries undertake assembly and trade costs are low enough that this assembly is undertaken for export as well as home markets. By contrast, countries in region IVB import all their final consumption of good *X*, supporting this by correspondingly larger export of capital intensive components. This corresponds to "vertical" investments, as different stages of the vertical production are undertaken in different countries. Countries in regions IVA and IVB are linked in production networks.

What do we learn from this analysis? First, the model encompasses production patterns that are often presented as alternatives deriving from different theoretical frameworks (see e.g. Markusen, 2002). In particular, it shows the coexistence of both horizontal and vertical investments; countries may engage in one or the other, horizontal being more likely in remote countries with "middling" factor endowments, while vertical is more likely in low trade cost countries with "extreme" factor endowments.

Second, the approach can be used to investigate the effect of economic changes, such as economic integration. For example, suppose that a single small labour abundant country reduces its trade costs—for example, an east European country joining the EU. Then this country moves down along the arrow *ab* in the Figure 1b. As it does this, its production structure changes with the growth of *X*-assembly for export. This is associated with a very large increase in trade volumes (components are imported and re-exported embodied in final assembled output) and with an increase in the price of the factor used intensively in assembly (labour).[3]

The results presented in this analysis are just a working through of comparative advantage theory. Our intuitions from comparative advantage remain useful, although the presence of several sources of comparative advantage (remoteness/ transport intensity as well as fac-

[2] For discussion of the concepts of horizontal and vertical foreign direct investment see Barba Navaretti and Venables (2004).

[3] This thought experiment was conducted changing trade costs for a single small economy at unchanged world prices. Of course, a more radical comparative static change would alter goods prices and shift boundaries between regions of the figure.

tor abundance/ factor intensity) together with the possibility of activities fragmenting, mean that there is an inevitable complexity.

1.2. Industrial location in the EU

The previous sub-section sketched out how we think a very long run situation—free of pre-EU trade barriers and pre-existing industrial structures—might be shaped. Of course, this is not what history has passed down to us, but how far away are we from this "long run" situation? This question has not been answered in a precise way, so argument usually proceeds by comparing the EU with an area of similar size that has not had a history of trade barriers and national champions—the US.

The comparison drawn by Krugman (1991b) was between the four large US regions (NE, MW, S and W) and the four largest countries in the EU. He showed that the industrial structures of the US regions were considerably more dissimilar from each other than were those of EU countries from each other. While it is difficult to make precise comparisons because of the inherently different sizes and geographies of the two areas, a somewhat finer comparison can be made by looking at the spatial concentration of a particular industrial sector relative to the spatial concentration of industry as a whole (the Hoover-Balassa index). Braunerhjelm et al. (2000) compute these measures for 8 broad sectors in the US and the EU; analysis has to be at the level of these very broad sectors to be comparable. In 6 of the 8 sectors production is more (relatively) spatially concentrated in the US than the EU, and the difference does not appear to be declining significantly through time. One of the other sectors is paper and pulp, determined largely by physical geography.

If there is less regional specialisation in the EU than in the US, is there any evidence of convergence? We can address this at a finer industrial level (26 industries) by computing a measure of how different each EU country's industrial structure is from that of the rest of the EU, and tracing the evolution of this measure through time.[4] To construct the measure of specialisation we calculate the share of industry k in country i's total manufacturing output (gross production value for each industry $x_i^k(t)$) and call this variable $v_i^k(t)$. Corresponding to this, we calculate the share of the same industry in the production of

[4] This section updates findings reported in Midelfart et al. (2002). A higher level of sectoral aggregation is required to include data for the last period.

all other EU countries, denoted $\bar{v}_i^k(t)$. We can then measure the difference between the industrial structure of country *i* and all other countries by taking the absolute values of the difference between these shares, summed over all industries,

$$K_i(t) = \sum_k abs\left(v_i^k(t) - \bar{v}_i^k(t)\right), \; v_i^k(t) = x_i^k(t) / \sum_k x_i^k(t),$$
$$\bar{v}_i^k(t) = \sum_{j \neq i} x_j^k(t) \Big/ \sum_k \sum x_j^k(t).$$

We call this the Krugman specialisation index (see Krugman, 1991b). It takes value zero if country *i* has an industrial structure identical to the rest of the EU, and takes maximum value two if it has no industries in common with the rest of the EU.

Table 1. Krugman specialisation index (production data, 4 year averages)

	1970-73	1980-83	1988-91	1994-97	1998-01
Austria	0.277	**0.252**	0.271	0.309	0.351
Belgium	**0.263**	0.296	0.318	0.383	0.437
Germany	0.304	**0.294**	0.345	0.352	0.375
Denmark	**0.523**	0.550	0.579	0.569	0.575
Spain	0.386	**0.266**	0.291	0.314	0.299
Finland	0.557	**0.471**	0.511	0.596	0.687
France	**0.122**	0.123	0.156	0.159	0.175
UK	0.195	**0.169**	0.190	0.180	0.227
Greece	**0.512**	0.557	0.626	0.709	0.744
Ireland	**0.679**	0.708	0.767	0.849	0.957
Italy	**0.333**	0.361	0.360	0.429	0.481
Netherlands	**0.479**	0.543	0.536	0.512	0.511
Portugal	0.524	**0.451**	0.559	0.557	0.608
Sweden	0.396	**0.389**	0.401	0.491	0.509
Average	0.396	**0.388**	0.422	0.458	0.495

Source: Author's calculations based on OECD data.

Values of these indices for each country are given in Table 1, computed for each of the 14 countries reported. They are calculated for four year averages[5] at the dates indicated, with bold indicating the minimum value attained by each country. The table reports them for each country and, in the bottom row, the average for all the 14 EU countries. Looking first at the averages, we see a fall between 1970/73 and 1980/83, indicating that locations became more similar. But from 1980/83 onwards there has been a more or less steady increase, indicating divergence. Turning to individual countries, we see that from 1970/73 to 1980/83 seven out of fourteen countries became less specialised, while between 1980/83 and 1998/01, all countries except the Netherlands experienced an increase in specialisation. That is, they became increasingly different from the rest of the EU.

The magnitude of the size of the changes is also informative. For example, given production in the rest of the EU, Sweden's specialisation index in 1998/01 took a value of 0.509, indicating that 25 per cent of total production would have to change industry to get in line with the rest of the EU (that is 50.9 per cent divided by 2, because the measure counts positive and negative deviations for all sectors). Thus, over the near 20-year span from 1980/83 to 1998/01 just 6 per cent of Sweden's production changed to industries out of line with the rest of Europe.

1.3. Why does specialisation proceed slowly?

From the analysis of the preceding section we see that EU countries are becoming more specialised, but that industry was, and remains, less spatially concentrated than it is in the US. Furthermore, the processes of structural change and specialisation seem to be relatively slow. Why is this so? Or more generally, what impediments are there to structural change?

There are a number of possible answers to this question. The first is that it might simply be the case that sunk-capital costs are substantial. While of some importance, this argument alone seems insufficient to account for the slowness of change, and there are few sectors where sunk costs are so large and capital so durable that this should support decades of persistence. However, policy can amplify these costs; in many EU countries there are substantial costs of plant clo-

[5] In order try to remove spurious fluctuations due to the differential timing of country and sector business cycles.

sure. Furthermore, other artificial barriers to economic reorganisation remain. It is clear that some EU governments continue to resist structural change through substantial use of state aids in support of ailing national champions, although Midelfart-Knarvik and Overman (2002) argue that this has had only limited effect on the direction of specialisation.

A variant on the sunk-capital costs argument is the existence of specialist skills in the labour force. These may be sector specific skills—engineering or financial services—or may be firm specific. Either way, if labour is immobile, it will be a force for inertia. Sutton (2000) argues that firms are best defined in terms of their "capabilities" and that these are typically embodied in their labour force. In some activities it is relatively easy to transfer these skills to workers in a new location by, for example, transferring a few key workers to train workers in the new location. In other activities specialist knowledge may be widely diffused between specialist workers. It may then be impossible to transfer knowledge by temporary transfer of a small number of key workers; essentially all workers would have to move. If this is not possible then the firm is locked into its existing location.

Further arguments come from the logic of agglomeration. Starting from the blank map, theory suggests that in industries prone to clustering there should be relatively few and large clusters of activity. But what happens starting from a situation in which there are already several clusters in place? Path dependency means that the outcome with few large clusters may not be reached.

To formalise this argument, suppose that agglomeration is due to cost and demand linkages, creating clusters of related industries, as in the automobile industry.[6] When trade costs are high there will be a number of clusters, e.g., each EU country will have its own automobile industry. As trade barriers fall what happens? Clustering is facilitated as the need to locate close to final consumers is reduced, and firms continue to get benefit from locating in a cluster with supplier and customer firms. However, small changes in trade barriers may produce no change in the equilibrium number of clusters; firms in a cluster derive profitability from the existence of the cluster, so are not induced to exit by marginal changes. There is therefore a story of "punctuated equilibrium". Trade barriers may continue to fall for some period of time and have no effect on the spatial organisation of

[6] This section draws on Fujita, Krugman and Venables (1999) Chapters 16 and 17.

production. A critical point is then reached at which there is discontinuous change; the old equilibrium becomes unstable and one or more of the previous clusters collapses, with production moving to remaining clusters.

This argument suggests that change may be slow. A further argument suggests that convergence to a comparison region (eg, the US) may be incomplete. For any given underlying parameters of the model, there are generally multiple equilibria. For example, the equilibrium might support n, $n-1$ or $n+1$ clusters, all of which are stable. The actual number observed is determined by the history of the economy. It is possible to construct examples in which a country with a long history of free internal trade (the US) has just one cluster of activity in a particular sector. Another, with a history of falling trade barriers (the EU) may have lost some clusters, but still retains two or three. Even if the regions have identical parameter values (trade costs and market size) the logic of clustering, multiple equilibria and path dependence does not necessarily imply identical outcomes across the two regions. History may mean that EU countries remain less specialised.

1.4. Industry location; conclusions

What do we learn from this analysis? Economic analysis seems to provide a reasonable interpretation of recent history, and changes are consistent with theory. However, it would be foolish to offer detailed predictions of future change. In addition to the inherent non-uniqueness that we have outlined there are important remaining unknowns, perhaps above all the role and nature of what we have called "disintegration costs". If a firm decides to outsource activities, spatially dividing some of its activities what costs does it incur? Direct monetary costs of shipment, transport etc. are conceptually straightforward and perhaps possible to quantify. Additionally there are time costs involved in shipment; delay slows down the speed with which firms can react to shocks to demand or costs, and this creates a significant barrier to fragmentation (Hummels, 2000; Harrigan and Venables, 2004). Other difficulties of management, of transferring firms' capabilities, and of implementing "just-in-time" production methods across a spatially fragmented production chain may be costly, but little is know about them. Work is needed in which economics engages with management and business expertise in identifying and quantifying these costs.

2. Income differentials

We now turn from industrial location to the distribution of per capita income across the EU. Of course, the two are tightly related. In equilibrium, income differences reflect either endowment differences, technological or institutional differences, or differences arising from geography. Studying the distribution of per capita income—or of individual earnings—is therefore of interest both in its own right and as an additional insight into the determinants of firms' location decisions.

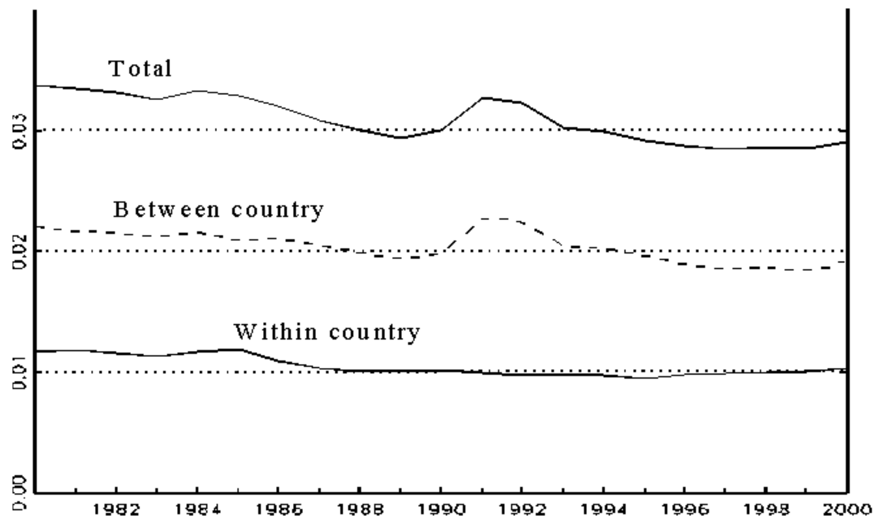
2.1. Centre and periphery

One of the key stylised facts about per capita income levels in Europe—and one of the key concerns of policy makers—is the existence of a "centre-periphery" wage gradient. A naïve regression of per capita income (across EU NUTS2 regions) on distance from Luxembourg yields a significant negative effect, such that each doubling of distance reduces per capita income by around 15 per cent. A rigorous theory based analysis of this relationship builds on Redding and Venables (2004) by constructing measures of the "market access" and "supplier access" of each of the NUTS2 regions, and using these as explanatory variables in the regression for per capita income, together with controls for educational and other differences between regions (see Breinlich, 2004). Such studies find significant market access and supplier access effects, yielding quantitative estimates of the role of distance from the centre of a similar magnitude; doubling distance reduces trade and hence market access and hence per capita income by around 15 per cent.

The pattern of convergence

Is there any evidence that these income gradients are flattening and wage gaps narrowing? There is an extensive literature on this subject that reaches the broad consensus view that there has been some narrowing of gaps between countries (as evidenced by the performance of Ireland, Spain and Portugal). However spatial inequalities within countries remains large, and in many countries (Italy, UK, Spain) have not diminished. The overall picture is given in Figure 2 which reports the time path of inequality in per capita income across 194 NUTS2 regions in 15 EU countries. A Theil index is used, and the upper line gives inequality between all the NUTS2 units. We see significant decline, despite an upwards blip in the early 1990's. The two lower lines decompose inequality into its between country (dashed line) and within country (lower solid line components). As is evident, between country inequality has declined much like total inequality. But within country inequality has been constant since the late 1980's, following a small decline in the mid-1980's.

Figure 2. Theil index of inequality, NUTS2 sub-regions



What is the bearing of geography on the spatial pattern of incomes across the EU? Table 2 presents some regressions that explore the determinants of the performance of different NUTS2 areas. The first two columns have as dependent variable per capita income in 1980-3. Independent variables are two simple measures of geography; distance from Luxembourg and density (population divided by area). Column 2 has fixed effects for each country, and column 1 does not. Without fixed effects we see that distance from Luxembourg has a significant negative effect—the income gradient referred to above—while density is insignificant. Adding country fixed effects means that coefficients are identified just from within country variation, and we find (column 2) that density has a positive and significant effect, while the coefficient on distance falls by a factor of three. Columns 3 and 4 repeat the same exercise for 1997-2000. Density coefficients increase in both equations, while there are very small reductions in the distance coefficients.[7]

Table 2. Regression results (geography and per capita income, NUTS2)

	1: Income pc 1980-83	2: Income pc 1980-83	3: Income pc 1997-00	4: Income pc 1997-00	5: Growth rate of income pc	6: Growth rate of income pc
Density	-0.0 (0.0)	0.077 (5.07)	0.025 (1.31)	0.091 (6.53)	0.025 (3.22)	0.029 (3.98)
Distance Luxembourg	-0.246 (-7.71)	-0.085 (-2.17)	-0.245 (-8.68)	-0.071 (-1.96)	-0.047 (-3.51)	-0.003 (-0.16)
Initial income pc					-0.195 (-7.43)	-0.206 (-5.99)
Country fixed effects	no	yes	no	yes	no	yes
No. of observations	194	194	194	194	194	194
R^2	0.26	0.75	0.34	0.74	0.26	0.61

Notes: All variables in logs: Results are based on OLS regressions: t-statistics in parentheses. Distance from Luxembough is actual distance in km plus 100 km; constant term omitted.

This suggests that within country variations in density are an important determinant of variations in per capita income. The evidence that density is becoming increasingly important is confirmed in columns 5 and 6. In these columns the dependent variable is the not the level of per capita in come, but its growth between 1980-83 and 1997-00. Independent variables are as before, plus the level of initial income in the region. As is usually the case, the level of initial income is significantly negative, indicating some underlying convergence. Distance from Luxembourg has a negative sign, indicating that—with the other controls of the equation present—there is no evidence of the income penalty of peripherality diminishing. Most interestingly, density has a significant positive effect, so growth has tended to be faster

[7] For detailed studies of the productivity effects of density that include a wide set of controls see Ciccone and Hall, (1996), Ciccone (2002), Rice and Venables (2004). The survey of Rosenthal and Strange (2004) suggests a range of estimates of the elasticity of productivity with respect to density of 0.04-0.11, compared to the estimates of 0.077-0.09 in Table 2.

in regions with high population densities. The second column adds country fixed effects, in order to focus on the determinants of differential growth performance within countries. The distance variable ceases to be significant, while the other two variables remain significant and with coefficients of broadly the same magnitude. This means that the density has an important positive effect on growth within countries, as well as between then.

Overall then, the picture painted by the spatial inequality measures and the regressions is one with the following three characteristics. (i) Some process of catch up by lagging countries and regions; (ii) little evidence of the centre-periphery income gradient systematically diminishing and; (iii) sub-national inequalities increasing somewhat due to the increasing importance of variations in the density of activity. Essentially, cities are doing relatively well, and it is to cities that we now turn.

3. The European city system

If dense city regions are doing relatively well, what implications does this have for the future economic geography of the EU? What happens if after product market integration and the adoption of the single currency, the EU achieves meaningful factor market integration? Increased labour mobility is likely to increase the extent to which agglomeration occurs, and possibly create pressure for expansion of some cities, perhaps at the expense of others. While it is very difficult to make predictions about the location of aggregate activity in a single Europe, in this section we speculate about possible developments by appealing to a well-known regularity that applies to city sizes.[8]

A feature of the urban system in many countries is that the city size distribution tends to follow the rank size rule, sometimes known as "Zipf's law". That is, if we rank cities by size from the largest to the smallest, then the *n*th city has population $1/n$ that of the largest.[9] Thus, the second largest city has population 1/2 that of the largest, the third largest city has population 1/3 that of the largest etc. This is illustrated in Figure 3, the left-hand panel of which plots the (natural) log of population against the (natural) log of the rank of city size for the top 100 US cities. Cities are ranked from largest to smallest. The

[8] This section draws on Midelfart-Knarvik et al. (2003).

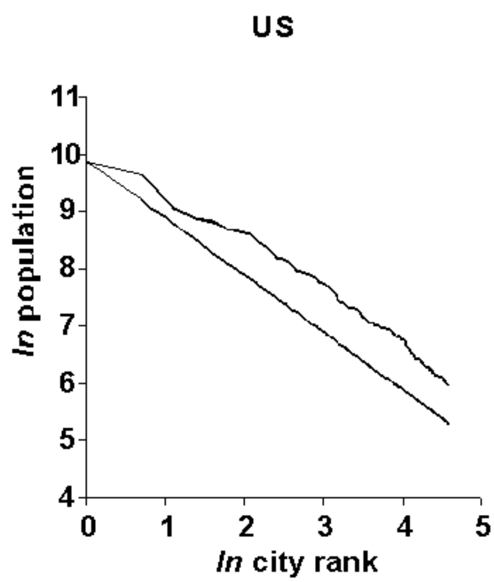
[9] Zipf's law states that the distribution of city sizes follows a Pareto distribution with coefficient equal to one. See Gabaix and Ioannides (2004) for more details.

highest ranked city is New York with a population of 19,876,488 in 1997.[10] The 100th largest city is Santa Barbara with a population of 198,760. The straight line shows what the graph would have looked like if US cities exactly followed the rank size rule. We can see that the US city size distribution is pretty close to obeying the rank size rule.

To make this statement more precise, we regress the log of population against the log rank of the city. For the top 100 US cities a simple OLS regression gives:

$$\textit{ln population} = 10.43 - 0.95\ \textit{ln rank}$$
$$(\text{s.e.} = 0.03)$$

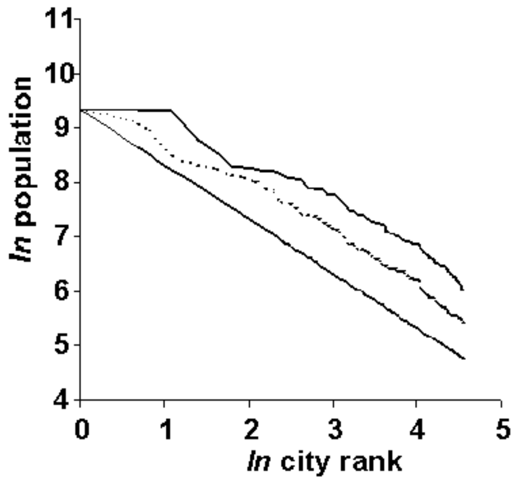
Figure 3. Zipf's law in the US and the EU



[10] Although we use the term "city", data are actually for Metropolitan Statistical Areas (MSAs) taken from the U.S. Bureau of the Census, State and Metropolitan Area Data Book 1997-98, Table B.1. For a variety of reasons, MSAs are the most appropriate spatial unit to take when considering the rank size rule for the US. See Cheshire (1999) for a discussion.

Figure 3. Continued....

EU25



The 95 per cent confidence interval for the zipf coefficient is (-0.88, -1.2) so we cannot reject the null hypothesis that Zipf's law holds.[11] The fact that the actual coefficient is less than one shows that US cities tend to be bigger than we would predict given their ranking relative to New York.

Figure 3 also shows the same graph for EU metropolitan areas.[12] The graph is plotted for the top 96 metropolitan areas in the expanded EU25. To give some idea of the underlying cities, the ten largest and ten smallest cities and their populations are listed in table 3. As before, the straight line shows what the graph would have

[11] As is well known, for the US, increasing the sample size up to approximately 140 cities brings the coefficient closer to 1. Increasing from 140 to 237 cities (the number of agglomerations with population larger than 50,000) moves the coefficient back away from one. See Black and Henderson (2003) for more details.

[12] Data on EU cities is taken from the world gazetteer (www.world-gazatteer.com). Data are for metropolitan areas with population greater than 400,000. Metropolitan areas may comprise several cities linked to one another economically, possibly extending across regional and national boundaries. Data comes from a variety of sources (official and unofficial) and so is not guaranteed to be strictly comparable across countries. Results are not sensitive to the inclusion or exclusion or particular cities or to fairly large changes in terms of the size of individual cities.

looked like if EU cities exactly followed the rank size rule. The upper line shows the true relationship for the EU and for comparison the dashed line shows what the city size distribution would have looked like if EU city sizes had the same relative sizes as in the US. That is, if the relative sizes of the Rhein-Ruhr (2nd) and Paris (1st), were the same as the relative sizes of Los Angeles (2nd) and New York (1st); the relative sizes of London (3rd) and Paris, the same as those of Chicago (3rd) and New York etc.

Table 3. Largest and smallest metropolitan areas in the EU25

Largest metropolitan areas			Smallest metropolitan areas		
Name	**Rank**	**Popul. ('000)**	**Name**	**Rank**	**Popul. ('000)**
Paris	1	11330.7	Grenoble	87	521.7
Rhein-Ruhr	2	11285.9	Szczecin	88	505.0
London	3	11219.0	Murcia	89	486.0
Ranstad (Netherlands)	4	6534.0	Belfast	90	484.8
Madrid	5	5130.0	Bari	91	480.7
Milano	6	4046.7	Montpellier	92	466.3
Berlin	7	3933.3	Bratislava	93	428.8
Barcelona	8	3899.2	Lublin	94	418.8
Napoli	9	3612.3	Messina	95	415.3
Manchester-Liverpool	10	3612.2	Coventry	96	409.1

From the figure it is clear that EU cities do not come as close to obeying the rank size rule as do US cities. Again, we can make the comparison more precise by considering a simple OLS regression for EU cities which takes the form,

$$\textit{ln population} = 10.05 - 0.82\ \textit{ln rank}$$
$$(\text{s.e.} = 0.04)$$

The 95 per cent confidence interval for the Zipf coefficient is (-0.74, -0.9), so we can clearly reject the null hypothesis that Zipf's law holds. Notice, further, that the coefficient that we estimated for the US (-0.95) does not fall within this confidence interval showing that the two coefficients are statistically significantly different from one another. As we work down the urban hierarchy in the EU, city sizes decrease much slower for the EU relative to both Zipf's law and the

US. That is, the EU urban population is much more dispersed than either of these benchmark cases.

This evidence suggests that the EU city size distribution varies markedly from that found for the US. In two recent papers, Gabaix (1999) and Duranton (2003) derive theoretical explanations for the emergence of Zipf's Law. Although the papers emphasise very different economic mechanisms (shocks to amenities and technological shocks respectively), they do share one common feature: Zipf's Law arises in integrated economies when labour is mobile. Thus, in these models labour mobility is a necessary condition for the emergence of Zipf's Law. This suggests the question, what might happen to the distribution of city sizes in the EU if (when) labour eventually does become mobile across member states?

Figure 3 already shows what would happen if we converged towards the US holding the size of the largest city constant around the 11 million mark. The relative city sizes of the top three cities would need to change so that the second and third city are both substantially smaller with populations of 8.9 million and 4.9 million respectively. This decline in city sizes would need to occur right across the urban hierarchy. If we take the US as an intermediate case, the smallest city we consider (Coventry) would see its population decrease from 409,100 to 229,700. 37 cities would shrink to populations below 400,000, and the total urban population in these cities would fall from 166.5 million to 102.6 million. If the distribution actually converged to the rank size rule, the second and third ranked cities would have populations around 5.7 and 3.7 million respectively, while Coventry's population would shrink to just 118,000. 67 other cities would see their population decrease below 400,000 and the total urban population more than halves to 58.3 million. Of course, these predictions of *falling* city sizes should not be taken too literally. The key point is that the *relative* size of the smaller cities will fall if the EU city size distribution converges to the rank-size rule.

Note that this prediction assumed that the largest city (be it Paris, Rhein-Ruhr or London) do not change in size. An alternative is to allow the size of the largest city to increase sufficiently so that the top 96 cities still accommodate the entire urban population that currently live in these cities.[13] Taking the rank size rule as a benchmark, this

[13] The urban population for any given system can be calculated using the area under the corresponding curve shown in Figure 3. Our first two examples took the intercept (the size of the largest city) as given and imposed a slope (relative city sizes)

would require the largest city to nearly triple in size to 32.4 million. At the same time, the second largest city would increase to 16.2 million and the fourth to eighth ranked cities would also increase in size. In contrast, the third largest city would shrink slightly to 10.8 million. All remaining cities would be smaller under the benchmark than they are in the current data. A more plausible scenario emerges if we impose relative US city sizes as a benchmark. Now, the populations of the largest and second largest city increase to 18.4 million and 14.4 million respectively. Again, the population of the third largest city shrinks somewhat to 8 million. Also, as before, other cities see their populations change. This time, the 4th to 10th ranked cities are bigger while all remaining cities see their population fall.

Of course, the actual numbers are of less interest than the overall trend. In all these scenarios the urban population becomes increasingly concentrated in just a few urban areas.[14] A number of comments are in order. First, there are likely to be efficiency gains from this process, as productivity benefits from agglomeration are realized; of course, these may be offset by adjustment costs, and also by congestion costs unless appropriate infrastructure is put in place. Second, it is important to combine this with an understanding of what cities of different sizes do. In the US, the very largest cities tend to be reasonably diversified, while medium to smaller size cities tend to be quite specialized (see Henderson, 1988, 1997, for details). This suggests an EU trend towards an outcome with some larger diversified cities and many smaller specialised cities.[15]

consistent with the rank-size rule while allowing the area under the curve (total urban population) to change. Our second two examples hold the area (total urban population) given and allow the intercept (the size of the largest city) to change as we impose the relevant slope (relative city sizes) consistent with the rank size rule.

[14] This statement on relative concentration would be true even if we allowed the total urban population to expand as the result of a higher urbanisation rate in the EU.

[15] The determinants of the relative numbers of diversified and specialised cities are not well understood and the relative numbers need not be constant over time. Duranton and Puga (2001) suggest that the existence of the two types of cities is intimately connected to the existence of product life cycles. In the early stages of producing a product firms locate in large diversified cities while they work out the best production strategy, only moving to specialised cities later in the product life cycle. Changes in the length of product life cycles could thus change the relative numbers of diversified and specialised cities.

As we have stressed, this exercise is highly speculative. However, both theory and the EU-US comparison indicate the likelihood of changes in urban structure in the direction outlined. Once again, more work is needed to think through these possibilities in greater detail.

4. Conclusions

We have painted a picture of slow European progress towards increasing specialisation and narrowing income differentials, coupled with strong performance of cities and possible pressure for change in the urban structure of the EU. This is all quite consistent with the predictions of economic analysis, although also suggests that a major change in the focus of this analysis is needed. Trade theory, and the economics of integration, has traditionally worked with countries as the geographical unit and manufacturing industries as the sectoral unit. Thinking about European integration suggests the need to look not just at countries, but at regions and cities. Manufacturing is now less than 15 per cent of national output, and to understand the success of particular regions and cities we need to develop a better understanding of comparative advantage in tradable services such as finance and the creative and knowledge based sectors.

References

Amiti, M. (1999), Specialization patterns in Europe, Weltwirtschaftliches Archiv, 135, 1-21.

Baldwin, R. E., Forslid, R., Haaland, J.I. and Midelfart-Knarvik, K.H. (2001), EU integration and outsiders: A simulation study of industrial location, in R.E. Baldwin and A. Brunetti (eds.) Economic Impact of EU Membership on Entrants. New Methods and Issues, Kluwer, Boston.

Barba Navaretti, G., Venables, A.J. et al. (2004), Multinational Firms in the World Economy, Princeton University Press, Princeton.

Braunerhjelm P., Faini, R., Norman, V.D., Ruane, F. and Seabright, P. (2000), Integration and the Regions of Europe: How the Right Policies Can Prevent Polarization. Monitoring European Integration 10, Centre for Economic Policy Research.

Black, D. and Henderson, V. (2003), Urban Evolution in the USA, Journal of Economic Geography 3, 343-372.

Breinlich, H. (2004), Economic Geography and the Regional Income Structure in the EU, Centre for Economic Performance, LSE, London.

Cheshire, P. (1999), Trends in sizes and structure of urban areas, in P. Cheshire and E.S. Mills (eds.), Handbook of Regional and Urban Economics, Vol. 3, North Holland.

Ciccone, A. (2002), Agglomeration effects in Europe, European Economic Review 46, 213-228.

Ciccone, A and Hall, R. (1996), Productivity and the density of economic activity, American Economic Review 86, 54-70.

Duranton, G. (2003), City size distributions as a consequence of the growth process, processed LSE.

Duranton, G. and Puga, D. (2001), Nursery cities: Urban diversity, process innovation, and the life-cycle of product, American Economic Review 91, 1454-1477.

Flam, H. (1992) Product markets and 1992: Full integration, large gains, Journal of Economic Perspectives 6, 7-30.

Fujita, M., Krugman P. and Venables, A.J. (1999), The Spatial Economy: Cities, Regions and International Trade, MIT Press, Cambridge MA.

Gabaix, Z. (1999), Zipf's law for cities: An explanation, Quarterly Journal of Economics CXIV , 739-767.

Gabaix, X. and Ioannides, Y. (2004), The evolution of city size distributions, in J.V. Henderson and J. Thisse (eds.), Handbook of Regional and Urban Economics, Vol. 4, North Holland.

Harrigan, J. and Venables, A.J. (2006), Timeliness, trade and agglomeration, Journal of Urban Economics, forthcoming.

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

Henderson, V. (1997), Medium size cities, Regional Science and Urban Economics 27, 583-612.

Hummels, D. (2000), Time as a trade barrier, Mimeo, Purdue University.

Krugman, P. (1991a), Increasing returns and economic geography, Journal of Political Economy 99, 483-499.

Krugman, P. (1991b), Geography and Trade, MIT Press, Cambridge, MA.

Markusen, J.R. (2002), Multinational Firms and the Theory of International Trade, MIT press, Cambridge, MA.

Markusen J.R. and Venables, A.J (2004), A multi-country model with multi-stage production and country-specific trade costs: A generalization of factor-proportions trade theory, Mimeo, University of Colorado, Boulder.

Marshall, A. (1890), Principles of Economics. 8th Edition, Macmillan, London.

Melitz, M.J. (2003), The impact of trade on intra-industry reallocations and aggregate industry productivity, Econometrica, 71, 1695-1725´.

Midelfart-Knarvik, K.H. and Overman, H.G. (2002), Delocation and European integration: Is structural spending justified, Economic Policy 35, 323-359.

Midelfart-Knarvik, K.H, Overman, H.G. and Venables, A.J. (2003), Monetary union and the economic geography of Europe, Journal of Common Market Studies 41, 847-868.

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

Rice, P.G and Venables, A.J. (2004), Spatial determinants of productivity: Analysis for the regions of Great Britain, CEPR Discussion Paper 4527, Centre for Economic Policy Research.

Rosenthal, S.S. and Strange, W.C. (2004), Evidence on the nature and sources of agglomeration economies, in V. Henderson and J. Thisse (eds.), Handbook of Regional And Urban Economics: Cities and Geography, Vol 4, North Holland.

Sutton, J, (2000), Rich trades, scarce capabilities: Industrial development revisited, STICERD Discussion Paper EI28, London School of Economics.

Venables, A.J. (1999) Fragmentation and multinational production, European Economic Review 43, 935-945.

Venables, A.J. (2003) Winners and losers from regional integration agreements, Economic Journal 113, 747-761.

Venables, A.J. (1996), Equilibrium locations of vertically linked industries, International Economic Review 37, 341-359.

Venables, A.J. and Limao, N. (2002), Geographical disadvantage: A Heckscher-Ohlin-von Thünen model of international specialisation, Journal of International Economics 58, 239-263.

Viner, J. (1950), The Customs Union Issue, Carnegie Endowment for International Peace, New York.