Economic Integration and Industrial Agglomeration

ANTHONY J. VENABLES*
London School of Economics

Abstract: This paper reviews recent research on industrial location, focusing on the way in which reducing barriers to trade may induce relocation of industry. Integration may cause industries to agglomerate in a few locations, this causing divergence of the structure of integrating economies, and possibly also divergence of income levels. Smaller locations will have lower real wages than large ones, although in the limit — as trade costs go to zero — factor price equalisation occurs.

I INTRODUCTION

How does economic integration affect the location of economic activity across space? Does integration lead to convergence of income levels across the integrating regions, or does it tend to pull activity into some regions at the expense of others? Attempts to answer this question have a long history. On the one hand, a neo-classical view of the world suggests that integration may encourage divergence of economic structure (as countries specialise according to comparative advantage), but convergence of income levels. The most extreme statement of this position is the factor price equalisation theorem, suggesting that completely free trade in goods will tend to equalise factor prices, even if factor mobility is not possible. On the other hand there is a substantial literature (associated with Perroux, 1955, Kaldor, 1972, Myrdal, 1957 and others) arguing that cumulative causation may take over, pulling activity into some regions at the expense of others, and thereby causing, or at least perpetuating, inter-regional inequality.

Recent research has brought this second position into mainstream economics. The work is based on formalising the “positive linkages” that may

exist between different agents in a particular location. These linkages may take the form of technological externalities — for example knowledge spillovers between firms. Or they may be "pecuniary externalities", in the presence of which expansion of one activity raises the profitability of others. If these linkages are strong enough then they can create positive feedbacks, whereby expanding an industry in one location raises the profits of firms at that location, encouraging further expansion and "cumulative causation". The outcome of such a process is the spatial agglomeration of activity, and consequent divergence of both economic structure and income across countries or regions.

In this paper I review some of the work that Paul Krugman and I have done (both separately and jointly) on this topic. The first element of research on the topic is the construction of a theory of the location of firms. This is based on firms in industries in which there are increasing returns to scale and imperfect competition. In addition to its empirical relevance, having increasing returns is necessary if an interesting theory of location is to be constructed; without increasing returns to scale firms do not face real location decisions, since they can, without any cost disadvantage, split into parts operating in all locations. Section II of this paper outlines a basic model of firm location, and reviews some of its implications.

The second element of the research is to explore the linkages between agents which may give rise to agglomeration. Our focus is on the pecuniary externalities that are created by imperfect competition. It turns out that these create agglomeration forces, the strength of which depend on levels of trade barriers between locations. Section III outlines these linkages, and Section IV demonstrates how integration — reductions in trade costs — may trigger a process of agglomeration, and thereby create inter-regional inequality.

Most of the research on agglomeration has, so far, concentrated on the way in which agglomeration can occur in economies which are identical in underlying structure. Section V of the paper presents a brief exploration of the way in which these agglomeration forces operate when economies are not identical, but differ in size. The tentative conclusion is that small countries may experience loss of industry in the early stages of integration, confirming the view that cumulative causation will draw industry towards "central" regions. However, the full implications of this remain to be worked out, and the paper concludes with an agenda for future research.

II THE LOCATION OF FIRMS

The point of departure is a theory of the location of firms in an imperfectly competitive industry. Each firm's location decision is based on essentially two
considerations. The first is the cost of inputs at each location (suitably adjusted for all input quality differentials). The second is the cost of market access — i.e., the cost of getting output to consumers. On the first of these there is little to say; other things being equal, firms go where inputs are cheaper. The second is more interesting, because market access costs interact with trade barriers in a way that is not immediately obvious.

Consider the following thought experiment. Suppose that there are two locations, each with the same costs, but with different market sizes. “Market size” refers to total expenditure on the product in the location, and there is a cost of shipping the product between locations. The location with the larger market size can be thought of as a “central” location, where these firms have access to many consumers at low trade costs, whereas the location with the smaller market size is “peripheral” — firms face trade costs in reaching many consumers.

For simplicity, suppose that location 1 has market size 1 and location 2 has market size 2. Suppose furthermore that location 1 has just one firm, and location 2 has two. The triples (a: b, b) in the body of Table 1 give the sales of each of these three firms, the first element being sales of the firm located in 1, and the second and third the sales of each of the firms located in 2. Columns of the table give sales in markets 1, 2, and total sales (final column).

Table 1: Firms’ Sales as Trade Costs Change

<table>
<thead>
<tr>
<th>Location 1; Market Size = 1</th>
<th>Location 2; Market Size = 2</th>
<th>Total Sales of Each Firm</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of Firms = 1</td>
<td>No. of Firms = 2</td>
<td></td>
</tr>
<tr>
<td>A (1: 0, 0)</td>
<td>(0: 1, 1)</td>
<td>(1: 1, 1)</td>
</tr>
<tr>
<td>I (1/2: 1/4, 1/4)</td>
<td>(2/5: 4/5, 4/5)</td>
<td>(18/20: 21/20, 21/20)</td>
</tr>
<tr>
<td>F (1/3: 1/3, 1/3)</td>
<td>(2/3: 2/3, 2/3)</td>
<td>(1: 1, 1)</td>
</tr>
</tbody>
</table>

Row A is autarky. Evidently the example is set up so that the total sales of each firm (equal, under autarky, to sales in the firm’s home market) are unity. Row F is completely free trade. Each firm then takes 1/3 of each market and, (bearing in mind that expenditure in markets 1 and 2 are equal to 1 and 2 respectively), sales are as described. The middle row, I, is constructed for an intermediate level of trade barriers. The level is set to be that at which each firm does exactly twice as well in its home market as does a foreign competitor. Elements in the triples have to add up to total expenditures in each market (1 in the first column and 2 in the second), so the numbers must be as illustrated in the middle row of the table. Adding the sales in markets 1 and 2 gives the total sales of firms.
The point to note from this example is the non-monotonicity of the total sales of firms in the two locations as integration occurs. In the initial stages of integration the firm in the small location suffers from the reduction in trade barriers (its sales go from 1 to 18/20), and firms in the large location gain (sales increasing from 1 to 21/20). In the later stages, firms in the large location suffer and the firm in the small location gains (as sales return to unity).

This numerical example makes the simple, but rather general point, that firms in small economies are particularly disadvantaged at intermediate levels of trade barriers. The intuition is that there are two opposing forces at work as transport costs fall. On the one hand, firms in the small economy are more dependent on foreign trade than are firms in the large economy, so gain relatively much from trade liberalisation. But on the other hand, there are more firms in the large economy than in the small, and each of these firms starts selling into the small market as trade costs come down. In the early stages of liberalisation the latter effect dominates (since initial trade volumes are zero), and firms in the small economy lose out to imports. In the latter stages this is reversed, and the benefits of being able to sell into the large market become relatively more important. This emerges as a matter of arithmetic in this example, but is a property of a much wider range of models (see Krugman and Venables, 1990).

The preceding example held the number of firms in each location constant. To study what happens to the location of industry we need to make the number of firms endogenous, and this can be done using a standard model of trade and imperfect competition. For example, consider a "Dixit-Stiglitz" model of monopolistic competition (Dixit and Stiglitz, 1977), which, in its multi-country form, has probably become the benchmark model of new trade theory (see for example Helpman and Krugman, 1985). The demand side of this model is characterised by product differentiation, separate product varieties each having iso-elastic demand curves, these curves being steeper the more differentiated are products. The supply side has each firm producing a single variety of differentiated product, and operating under increasing returns to scale. There is imperfect competition, with price cost mark-ups determined by the slopes of demand curves, and the equilibrium number of firms determined by the condition that each firm should make zero profits. The multi-country variant of this has the location of industry (i.e., the number of firms operating in each country) determined by zero profit conditions for each country.

If we restrict ourselves to looking at just two locations (1 and 2), then the equilibrium location of industry can be summarised in the following way. Let N denote the relative location of the industry, that is the number of firms in
location 1 divided by the number in location 2. C denotes costs in location 1 relative to 2, and S denotes country 1 expenditure on the product relative to that in country 2. t is the proportional trade cost factor, so t = 1 is completely free trade, and t = 1.5 means that trade costs amount to 50 per cent of the value of output; these trade costs measure all the costs of shipping between locations—tariffs, non-tariff barriers, transport costs, and any costs imposed by language or institutional differences. The equilibrium location of industry is a function of relative costs, expenditures, and trade costs, and we shall summarise this relationship by the function f, so

\[ N = f(C, S, t) \]  \hspace{1cm} (1)

As expected, higher relative costs mean fewer firms in location 1 relative to location 2; greater expenditure means more firms; and trade costs can, as we have already seen, operate in either direction.

The function f is illustrated on Figure 1. The vertical axis is C, the level of production costs in location 1 relative to 2, the horizontal axis measures trade costs, t, and the figure is drawn for the case of S = 0.1 — i.e., country 1 having a much smaller market than country 2. The curves are iso-N lines. Thus N = S is the combination of t and C at which industry is divided between locations in proportion to market size, and so on. Clearly, if N/S < 1, then the small country is a net importer of the good.

The effects of economic integration on the location of industry at unchanged relative costs are found by moving horizontally from left to right across this figure. We see that if C = 1 then reducing trade costs brings steady relocation of industry from the small location to the large, with the small location losing all industry when the N = 0 contour is passed. However, in the limit when trade costs are zero (t = 1) then firms are indifferent about their location.

If the small economy has a cost advantage (say C = 0.95), then the picture that emerges as trade costs are reduced is one of relocation of industry from the small economy to the large, until trade costs are quite low (in the figure, around 1.2) followed by a reversal of the process. The intuition here is that in the initial stages of integration the market access forces dominate, pulling industry to the location with the larger market. But at low enough trade costs, industry becomes more "footloose" and more sensitive to production cost differences; industry then moves to the lower cost location.

What happens if relative production costs, C, are made endogenous, rather than exogenous, as would be the case if changing industrial location affects factor prices in each country? This can be seen most easily by going to an
extreme case where the industry under consideration uses a specific factor, and the relative endowment of the factor is equal to relative market size. Full employment of the specific factor then requires that $N = S$, and the price of the factor must therefore move so that relative costs, $C$, follow the $N = S$ locus in Figure 1. As is clear from the diagram, at high trade costs relative production costs (and relative wages) must decline for the small economy. At lower trade costs this is reversed, as the disadvantage of the small economy is diminished, and industrial location becomes increasingly sensitive to production cost differences. As $t$ goes to unity (completely costless trade), so the $N = S$ locus converges smoothly to the limiting value implied by factor price equalisation. The information presented in Figure 1 gives relative costs (or wages) measured in terms of some numeraire good. However, the message of the figure carries over to real wages. The small economy experiences a reduction in real wages in the early stages of integration, followed by real wage increase. The large economy experiences real wage growth in early stages of integration, and either slower growth or possibly real wage decline as trade costs become very small.

The analysis of this section teaches us two things. First, the effects of integration on small or “peripheral” economies are ambiguous. At inter-

![Figure 1](image-url)
mediate trade costs the small location is disadvantaged, and this can manifest itself in different ways; reduced sales per firm, if the number of firms is held constant; exit of firms, if wages are held constant; or a reduction in wages, if labour demand impacts on wage rates. The second message is that the location of demand may have a disproportionate effect on the location of industry, in that locations with small demand are net importers of the product. This provides a basis for the "demand linkage" mechanism which we shall see in the next section.

III AGGLOMERATION FORCES

The story so far shows that centre/periphery distinctions are of importance in assessing the effects of economic integration, but do not capture "cumulative causation" in the development of regional inequalities. To do this we need to add "positive feedbacks" or "linkages" between the actions of various decision takers in the system. Before discussing what these linkages might be, consider Figures 2a and 2b. The horizontal axis has the relative number of firms in the locations, N, and the vertical has the difference between the equilibrium and the actual value of N, i.e., it measures \( f(C, S, t) - N \). Figure 2a captures the case we discussed in the previous section. If relative costs, C, and relative market sizes, S, are independent of N, then \( f - N \) has a 45° negative gradient, as illustrated. Out of equilibrium dynamics are illustrated by the arrows lying on the horizontal axis. If \( f \) is greater than \( N \), then profits are positive and entry occurs, and if \( f \) is less than \( N \) then there is exit, giving the directions of motion illustrated. Clearly the equilibrium point is at \( E \), and this equilibrium is stable.

If relative costs and/or market size are functions of relative numbers of firms then the gap between equilibrium and actual relative location can be written as \( f(C(N), S(N), t) - N \). Figure 2b illustrates a case where these effects operate to make \( f \), the equilibrium number of firms an increasing function of the actual number, \( N \), and, furthermore, the effects are so strong that the gradient of \( f - N \) goes from negative to positive. Our dynamics give entry when \( f > N \), this now being to the right of point \( E \). \( E \) is therefore an equilibrium, but it is unstable. For the case drawn, there are two stable equilibria. One is at point \( A \) with \( N = 0 \), and the other simply has the country labels reversed, so is at \( 1/N = 0 \). In each of these cases all firms have agglomerated in a single location, and entry in the other location is unprofitable.

This simple picture illustrates how changing the slope of the function \( f(C(N), S(N), t) \) may destabilise the diversified equilibrium and create agglomeration. But what are the "linkages" generating dependence of costs
Figure 2a.

Figure 2b.
and market size on the number of firms in each location, and can these be large enough to reverse the gradient of $f - N$?

Krugman (1991a,b) demonstrates that labour mobility can serve to create linkages and hence agglomeration. As we saw, the model outlined in Section II has the property that the larger region has higher real wages than does the small region. Evidently, if labour is mobile, it will flow from the small region to the large. As labour moves it takes its spending with it, so the large market gets larger and the small smaller, amplifying the wage difference, and possibly leading to agglomeration. (Supplies of other factors will stop the economy from emptying all together.) We can relate this to Figures 2a and 2b in the following way. In the Krugman model, it is as if when firms move they take workers with them. Relative market size, $S$, is therefore an increasing function of the relative number of firms (and workers) $N$. We shall call this a "demand linkage" — it says that the more firms there are in a location the higher is demand for firms' output at the location. If this demand linkage is powerful enough then it will give configuration 2b, with agglomeration, rather than configuration 2a. Evidence in the US (for example Blanchard and Katz, 1992) suggests that migration may be significant enough to create these forces for US states, although migration in Europe is perhaps insufficiently responsive to economic factors for this mechanism to be of much relevance to European integration.

An alternative possibility is that there is a "cost linkage", $C(N)$. One form this could take is positive technological externalities that are of limited geographical reach. In this case $C$ is a decreasing function of $N$ and (since $f$ is decreasing in $N$) this can, if powerful enough, give rise to agglomeration. It is not very satisfactory to simply assume the presence of such externalities; there must be some transmission mechanisms which creates these spillovers, and it would be nice to see this modelled explicitly. Nevertheless, it seems clear that knowledge spillovers of some sort may well be important in certain industries — high technology and Silicon Valley spring to mind.

A third possible mechanism arises if there is an input-output structure between firms in imperfectly competitive industries (Venables, 1993; Krugman and Venables, 1993, 1994). Although it is natural to think of this in a multi-industry setting, we can see how it operates in our framework by supposing that industry uses some of its own output as input. This will generate both demand and cost linkages. The more firms there are at a location then the larger will be demand, as firms are demanding the industry output as an intermediate input; we therefore have $S(N)$ an increasing function. And furthermore, the more firms there are at a location the lower will be the cost of intermediate goods because, if there are transport costs on shipping varieties of intermediates, fewer intermediates bear transport costs;
this means that $C(N)$ is a decreasing function. Both these effects work in the direction of agglomeration, as in Figure 2b.

IV INTEGRATION AND AGGLOMERATION

We argued in the preceding section that the interaction between imperfect competition and intermediate goods create demand and cost linkages between firms. Whether or not these are powerful enough to cause agglomeration depends on a number of parameters. These include the magnitude of the input-output linkages, the degree of imperfect competition, and the level of trade costs between regions. It is the role of trade costs that makes these effects important in the study of economic integration.

Possibilities are illustrated on Figures 3a, 3b and 3c. All three are constructed with the same input-output linkages and the same price-cost mark-ups, but with different levels of trade costs. Figure 3a is drawn with high trade costs, this giving an equilibrium at E with production diversified between the locations. The intuition for this is clear. At very high trade costs (close to autarky), the presence of consumer demand in each location means that the industry has to produce in both locations. Figure 3c is drawn for a low level of trade costs. This reduces the need to be close to final consumption, making forces for agglomeration relatively more powerful. The equilib-

![Figure 3a.](image-url)
Figure 3b.

Figure 3c.
rium at E is unstable, there is a stable equilibrium at A, and a further stable equilibrium at plus infinity (1/N = 0). The case of intermediate trade costs is illustrated in Figure 3b. The diversified equilibrium at E is stable, but so too is agglomeration — the equilibria at N = 0 and 1/N = 0. In between these stable equilibria are unstable equilibria, at the points B.

The dependence of the structure of equilibrium on trade costs is drawn out more explicitly in Figure 4 which has trade costs on the horizontal axis, and the absolute numbers of firms in each location, denoted n₁ and n₂, on the vertical. The numbers of firms are computed from a numerical example. Solid lines on the figure represent stable equilibria, and dashed lines unstable equilibria.

The example illustrated in the figure is constructed with the two economies identical in underlying preferences, technology, and endowments, implying that the diversified equilibrium has equal division of industry between the locations, n₁ = n₂. At high levels of trade costs (t > 2.5) diversification is the unique (and therefore stable) equilibrium. With trade costs in the region 1.85-2.5 there are five equilibria. The symmetric diversified case (n₁ = n₂), the equilibrium with industry entirely in location 2 (as illustrated), the equilibrium with industry entirely in location 1 (not illustrated) and two unstable equilibria. Below 1.85 there are two stable agglomerated equilibria (the one illustrated has industry entirely in location 2), and the diversified outcome remains an equilibrium, although it is unstable. Economic integration moves us from left to right on this diagram. Starting with a stable diversified
equilibrium there is a critical point (point B, at approximately \( t = 1.85 \)) at which the diversified equilibrium is rendered unstable and industry relocates, agglomerating in a single location. However, there is nothing in the theory that says whether the agglomeration occurs in location 1 or in location 2.

The figure is constructed under the assumption that wages — and the prices of any other primary factors used in the industry — are held constant and equal. Under these conditions it is possible to show that the structure of equilibria must be qualitatively as illustrated (Venables, 1993). That is, providing there is both imperfect competition and an input-output linkage, then it must be the case that the system has a critical point (point B, at which the diversified equilibrium flips from being stable to unstable), at some positive level of trade costs, \( (t > 1) \). This critical value is greater (and hence the range of trade costs in which agglomeration occurs is larger) the higher is the price-marginal cost mark-up generated by imperfect competition, and the greater is the share of intermediate goods in production.

The mechanism driving this agglomeration of activity is the interaction between increasing returns to scale and input-output linkages. At what level of industrial aggregation should we expect these forces to be most powerful? Do they operate at the level of particular industries or groups of industries which are relatively tightly linked by input-output connections? Or are the linkages important for manufacturing activity, as a whole?

The former case is studied in Krugman and Venables (1993). At high trade barriers each industry operates in each location, but as trade barriers come down, so there is a process of agglomeration in which some industries concentrate in one location, and others in other locations. Applying this to Europe suggests that the economic geography of Europe may become more like that of the US. Regions — or countries — lose their presence in some industries, and industries become more geographically concentrated. However, each region or country may have some cluster of industries so that although there is divergence of the structure of economies, there need be no divergence of income. The welfare economics of this case are straightforward. Gains from integration are particularly large (as industries reap benefits from agglomeration) and may be quite evenly distributed across regions — each region has labour demand from its cluster of industries. However, integration may well give rise to significant adjustment costs as industries relocate, countries lose a presence in some industries, and perhaps also certain types of skills become redundant in some locations.

The case in which there are strong linkages through manufacturing as a whole is studied in Krugman and Venables (1994). In this case agglomeration forces will tend to make all manufacturing locate in the same place.
Evidently, we can no longer continue to assume that wages are exogenous, but must now allow them to respond to potentially widely different levels of labour demand. Linking wage rates to industrial labour demand modifies analysis in two ways.

First, if relative wages depend on the location of industry, then this will tend to offset agglomeration forces, since it will make C(N) an increasing not a decreasing function. It is possible that this effect may be large enough to prevent agglomeration from occurring. (In the extreme case in which each location has a fixed labour supply to industry, then obviously agglomeration cannot occur — industry has to go where the labour is). An interesting case arises if the wage effect is less powerful than this. The outcome will then look as it does in Figure 4 until quite low trade costs are reached. However, at zero trade costs it must be the case that the diversified equilibrium with \( n_1 = n_2 \) is unique and therefore stable. This is because agglomeration effects are absent (the location of intermediate goods suppliers and final demand is completely immaterial when there are no trade costs). Industry therefore goes to whichever location has the lower wage costs, which means that in equilibrium wages must be equal, implying equal labour demands and equal division of the industry. The overall pattern of equilibria as trade costs go upwards from free trade is therefore as follows; at very low trade costs there is a unique equilibrium, with factor supply considerations dividing production between locations; at somewhat higher trade costs there are five equilibria — stable equilibria with and without agglomeration plus two unstable equilibria; at higher trade costs again, the pattern is as in Figure 4, with three, then five, and then one equilibrium. Essentially then, at very low trade costs location of industry is determined by factor supply considerations; at very high trade costs by final consumer demand; and at intermediate levels of trade costs, agglomeration forces are dominant.

Endogenous wages not only dampen the incentive to agglomerate, but also mean that when agglomeration does occur it is associated with wage differentials between locations. The location with the agglomeration has higher labour demand and higher wages (although the wage differential is not so large as to outweigh the benefits of agglomeration). In this case integration may cause not only divergence of economic structure, but also divergence of real incomes. This point can be put more strikingly. Suppose that there is a reduction in trade costs which triggers agglomeration. Then the location that loses industry suffers a real wage decline, and the location that gains industry a real wage gain. The implications of this are drawn out more fully in Krugman and Venables (1994).
V INTEGRATION, AGGLOMERATION AND COUNTRY SIZE

The preceding section looked at possible effects of integration under the assumption that the underlying structure of the integrating economies is symmetrical. What happens when this assumption is removed? Such asymmetric cases have not yet been worked out in any detail. In this section we merely illustrate some possibilities for a case in which the two economies are of different size although — by assumption — have equal and constant wages.

The story is given in Figure 5. The construction of this figure is similar to that of Figure 4, except that country 2 is now assumed to be twice the size of country 1. At high trade costs the equilibrium is diversified, although there are more firms in 2 than in 1, as would be expected. Reducing trade costs amplifies the difference between economic structure, rather as we saw in Section II. At first there is continuous response to lower trade costs, but at some critical value, B, the agglomeration forces become relatively powerful enough to make the diversified equilibrium unstable, so the system flips to agglomeration. However, whereas in Figure 4 there were multiple equilibria, it now has to be the case that (if country 2 is sufficiently larger than 1) there is a unique equilibrium. The agglomeration must occur in the larger country, so it is certainly the case that \( n_2 > 0 \) and \( n_1 = 0 \).

Figure 5 describes a situation in which wages in the two countries are constant and equal. We can conjecture about the effects of removing this assumption and letting wages adjust in response to industrial labour.
demand. It then seems likely that integration will produce a pattern of wage divergence followed by convergence. As in Section II, the small economy will suffer declining wages to some point, followed thereafter by increasing wages and, in the limit, factor price equalisation. However, it is certainly going to be the case that the presence of intermediate inputs and the associated forces for agglomeration will amplify inter-locational differences, creating larger inter-locational wage differences than we saw in Section II.

VI CONCLUSION

Imperfect competition based models of intra-industry trade have become widely accepted as appropriate frameworks within which to think about a large part of world trade. The research outlined in this paper makes one modification to a standard model of this type; it allows for input-output linkages between firms in imperfectly competitive industries. Making this apparently minor modification fundamentally changes the structure of equilibria generated by the model, creating the possibility of “positive feedbacks” which support agglomeration of economic activity.

The work so far undertaken suggests many directions, both empirical and theoretical, for future research. On the empirical side, is there evidence that integration has fostered geographical concentration of industry? Some evidence, (both time series and cross country comparison) suggests this may be so, but work needs to be done to disentangle agglomeration effects from the ordinary process of specialisation predicted by any trade theory. If linkages exist, how strong are they? If a German car component manufacturer or City of London financial institution were to relocate to Ireland — or to India — how low would wages in these locations have to be to offset agglomeration advantages? Perhaps case study methods can shed some light on these issues.

On the theoretical side, much work remains to be done on incorporating locational asymmetries into the models, and on developing dynamics. Perhaps most fundamentally, the policy implications of this sort of model need to be developed. The implications of this line of research for regional policy remains uncharted territory.

REFERENCES


