SWEDISH ECONOMIC POLICY REVIEW 12 (2005) 9-27

Industry location: The causes

Richard E. Baldwin*

Summary

■ The location of industry is determined by the complex interaction of many factors, so it is useful to abstract from reality and focus on the main forces. In this paper I suggest it is useful to organise the various causes into three main categories. The first concerns physical geography, so-called first-nature geography. The second is the balance of economic agglomeration forces and dispersion forces—so-called second-nature geography. Most of these causes can be manipulated by policies such as production subsidies, trade liberalisation, and taxation. I also suggest that there is an important "in between" category, namely causes that adjust more slowly than industrial clusters but faster than coastlines. In this "1.5 geography" I would include transport networks and factor endowments, both of which are malleable to government policy. ■

Key words: H32, P16.
JEL Classification: Industrial location, first nature geography, second nature geography, 1.5 nature geography.

* *Richard E Baldwin is Professor at Graduate Institute of International Studies, Geneva.*

SWEDISH ECONOMIC POLICY REVIEW 12 (2005) 9-27

Industry location: The causes

Richard E. Baldwin*

1. Introduction

Spatially speaking, the economy is a very lumpy place at almost any level of resolution—continents, nations, provinces, cities, or neighbourhoods. Some of this bunching is trivia—oil extraction clusters in Saudi Arabia and logging clusters in Canada—yet much of the geographic clustering, especially that of industry, seems to be supported by agglomeration economies.

I was first introduced to agglomeration economies at a 1993 conference that brought together academics and government practitioners on the subject. One of the practitioners—a European Commission official in charge of Spanish structural spending—explained how he spent the billions of euros and what he hoped this would achieve. Importantly, he did this with no reference to economic factors that caused the poor regions to be poor. During the Q&A after his talk, one of the academics—Paul Krugman to be specific—asked him: "Why do you think these regions are poor?" The official responded that they were poor since few new firms invested in the regions and the existing firms were shutting down or moving out. On top of that, the talented young people tended to leave the regions as soon as they could.

Krugman realised that this was not an answer, but rather an elaborate restatement of the question. A lack of investment and a brain-drain are not causes of poverty; they are the symptoms of it. Wishing to be polite, Krugman gently pushed the official. "But, why do you think firms and young people leave?" Krugman prodded. The official stared at him with amazement, wondering how this professor could be so naïve. "Well," he replied, "because these are poor regions."

At first blush, this story makes the official out to be an intellectual lightweight. Upon reflection, however, one can see that this parable gets at a deep and difficult truth of industrial location. The location of

* *I would like to thank Karolina Ekholm, Tony Venables, and Philippe Martin for comments and suggestions.*

industry inevitably involves a chicken-and-the-egg causality. Firms tend to locate in big markets since big markets provide them with easy access to everything they need on the supply side and on the demand side. However, in locating to the big markets, the firms make the big markets bigger. The technical phrase for this circular causality is agglomeration forces.

1.1. Towards an organising framework

There are a number of ways to organise our thinking about the causes of industrial location, but I believe the most useful is a modification of the distinction between so-called first-nature and second-nature features. The economic geography literature distinguishes regional concentration of economic activity that can be attributed to physical geography features on one hand (first nature) and agglomeration economies attributed to the interaction of economic agents on the other hand (second nature).

This twofold cataloguing of forces, however, is really too blunt. Very little of the earth's geography is truly fixed. Much of Holland used to be under water, the Atlantic and Pacific Oceans are joined by a canal, a dozen tunnels pierce the Alps, and there is a bridge over the water barrier between Sweden and Denmark. Nevertheless, it is extremely useful to start by taking physical geography as given, since man finds it worthwhile to alter first nature geography only when second nature geography renders it profitable. At the other extreme, the location of a particular factory or even a whole sector can be quite unfixed. Fairly small changes in tax policy, or shifts in comparative advantage may induce an industry or firm to shift its location to an entirely different nation or even continent.

At a deeper level, the first nature/second nature distinction is really about the fixity of features, or, to use economic jargon, about the exogeneity of features. From this perspective, surely there is a continuum of fixity of features that "cause" industry location. In between these extremes are a series of quasi-fixed factors. For example, the spatial distribution of Europe's population is quite stable since most Europeans prefer not to move. This means that the geographic allocation of human resources is fairly fixed, at least at the national and maybe even sub-national level. However, over a period of decades, the composition of a nation's factor endowment can change in a major way in response to policy initiatives and agglomeration forces.

Table 1. Economic activity and welfare measures (EU27; 1998)

Regions:	Land share (percent)	Population share (percent)	GDP share (percent)	Unemployment rate relative to EU27 average (EU27=100)	Youth unemployment rate relative to EU27 average (EU27=100)	Share of Population with income above EU27 average (percent)
Core	14	33	47	74	61	89
Intermediate	21	26	32	101	95	70
Peripheral	65	41	21	121	134	18

Source: Baldwin and Wyplosz (2003) based on European Commission data.
Notes: EU27 includes the EU15 and the 10 nations that joined in 2004, plus Bulgaria and Romania. Regions are defined at the NUTS2 level of aggregation; see http://europa.eu.int/comm/regional_policy/ for definitions.

.

One quasi-fixed factor that perhaps deserves its own moniker—call it first and a half nature—is what might be called the economic infrastructure. This consists of some physical features, like the transport network, that are physically difficult to alter quickly, and some, such as knowledge networks that are not physically difficult to alter but nevertheless play an important role in sustaining the massively unbalanced spatial distribution of economic activity that we observe in industrialised nations. These features of the economic landscape are far from fixed, but are also extremely expensive to change rapidly.

In what follows, I cover the basic logic of how industrial location is affected by first, second and 1.5 nature geography. Before turning to the logic, however, it is important to ask why we care about industrial location in the first place. Why should anyone care about the lumpiness of economic activity? Few people live in the Alps, so why should it be a problem that there is so little economic activity there?

The answer, I believe, stems from the fact that a low density of economic activity tends to be associated with many of the worst evils of the social market economy, as Table 1 shows. For example, the periphery's unemployment rate is much higher than it is in the core, especially among young workers (34 percent above the EU27 average in the periphery compared to 39 percent below the EU average in the core). Only 20 percent of the people located in the periphery have above-average incomes while the figure for core-based people is almost 90 percent (European Commission, 2003).

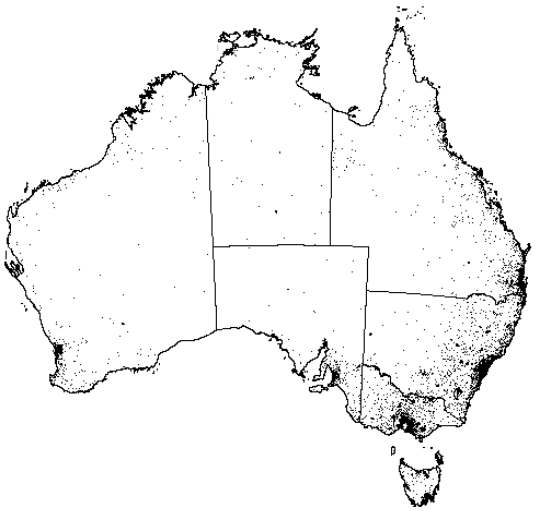
2. First and second nature geography

First nature geography is the obvious part of industrial location. Industries tend to cluster near ports and rivers since this allows them to obtain inputs and ship outputs at a reduced cost. This point manifests itself quite clearly in Figure 1. which shows the population on the Australian continent.

First nature geography is also the part of the industrial location that modern governments have to take pretty much for granted, although there are a few spectacular example of governments who have altered physical geography in order to encourage economic activity. The English Channel tunnel and the Kansai airport are two examples of this.

Second nature geography is more complex.

Figure 1. Australian population distribution



Source: Starr et al. (2004).
Note: One dot = 1000 people.

2.1. New economic geography

Much of the geographic grouping of production that is so prevalent in modern economies is supported by agglomeration economies where these are defined as the tendency of a spatial concentration of economic activity to create economic conditions that foster the spatial concentration of economic activity. This assertion, however, is incomplete.

Explaining industrial clusters with agglomeration economies is both trivial and baffling. Trivial since its very definition shows that assuming agglomeration economies is tantamount to assuming the result. Baffling since it is hard to know how a clear-headed theorist should approach this seemingly self-referential problem. The chief concern of the so-called new economic geography has been to open up the "black box" so that we can examine the economic logic driving self-reinforcing spatial concentration.

2.2. Agglomeration and the new economic geography

A combination of scale economies and trade costs generates forces that encourage geographic clustering of economic activity. This clustering can take two distinct forms:

- Overall clustering that results in some areas with lots of economic activity and some areas with almost none.
- Sectoral clustering where each sector clusters together in a region, but different sectors cluster in different regions, so all regions end up having some industry.

There are many agglomeration forces, and many schemes for categorising them. I will focus on the distinction between agglomeration forces that work though factors of production and those that work through goods.

Agglomeration forces and productive factors

There are two main stories on the factor side. The first is the classic Marshallian spillovers story. For reasons that are not entirely clear, some types of capital and labour are more productive when they work in spatial clusters. Silicon Valley and the City of London are the classic examples. Despite astronomical land and housing prices, hi tech firms find it profitable to set up in Silicon Valley. The story is that the knowledge spillovers that occur in the valley are more than enough to make up for the high costs. Similar stories are told for why banks from around the globe set up office in the priciest district of one of the world's priciest cities. Plainly, this is very close to assuming the result since the economic channels for these spillovers are not clear. Moreover, the spatial scale on which these externalities work is also not clear. Many of them only operate on a very local scale.

The second main story concerns labour market pooling. That is, firms are never sure which types and how much labour they will need, so they find it advantageous to locate near a big, thick labour market. But of course, the presence of so many firms attracts lots of workers, so the fundamentally circular nature of agglomeration force is in action.

At the level of regions and nations, agglomeration forces that operate via factors are of limited interest. The goods-channel, by contrast, has not natural spatial boundary. Since goods are sold around the world, agglomeration forces that operate via goods can concern

large scales spatially. The two most important agglomeration forces that operate via goods and thus across great geographical spaces are called demand linkages and cost linkages.

Agglomeration forces and goods: Backward and forward linkages

To illustrate the logic of these backward and forward linkages, it helps to make some bold assumptions. First, we assume that firms must choose one location, rather than, for example, producing a little bit everywhere. Second we assume that there are only two possible locations, a region called "north" and a region called "south".

The demand linkage rests on market size issues. Firms want to locate where they will have good access to a large market in order to reduce trade costs. This is where demand linkages start. Firms want to be in the big market, but in moving to the big market, they tend to make the big market bigger. For example, the firms directly affect market size since firms buy goods from each other. Firms also affect the market size indirectly since workers move to be near their jobs and they spend their salaries locally. This is an agglomeration force since spatial concentration of economic activity creates forces that encourage further spatial concentration.

Cost linkages work in a similar fashion. The difference is that it involves the cost of production rather than the cost of supplying consumers. Firms buy inputs such as raw materials, intermediate goods, machinery and equipment as well as specialised services such as marketing and financial services. Due to trade costs and other distance-related costs such as information costs, these inputs tend to be cheaper in locations where there are lots of firms supplying these inputs. Thus the cost linkage works by encouraging firms to locate near their suppliers, but since firms also supply other firms, moving to a low-cost location for intermediates tends to lower the cost of intermediates in that location even further. Again, this involves the classic circularity of an agglomeration force.

Dispersion forces

There are, of course, many forces opposing concentration and these are called dispersion forces. For example, land prices and the cost of some forms of labour tend to be higher in built-up areas. This counteracts the agglomeration forces by increasing the attractiveness of less developed regions.

While these congestion-based dispersion forces are important in the real world, they greatly complicate the analysis. Thus when we start to layout the interplay of agglomeration and dispersion forces, we ignore them to begin with. Indeed, the sole dispersion force we consider is the so-called local competition force. That is, given trade costs and imperfect competition, firms are naturally attracted to markets where they would face few locally based competitors. In seeking to avoid local competition, firms spread themselves evenly across markets, which is why we call this a dispersion force.

Of course, the pro-concentration (agglomeration) forces and anti-concentration (dispersion) forces operate simultaneously and the equilibrium outcome is a geographic distribution of economic activity. The key question is how economic integration affects the equilibrium location of industry.

2.3. The EE-KK diagram

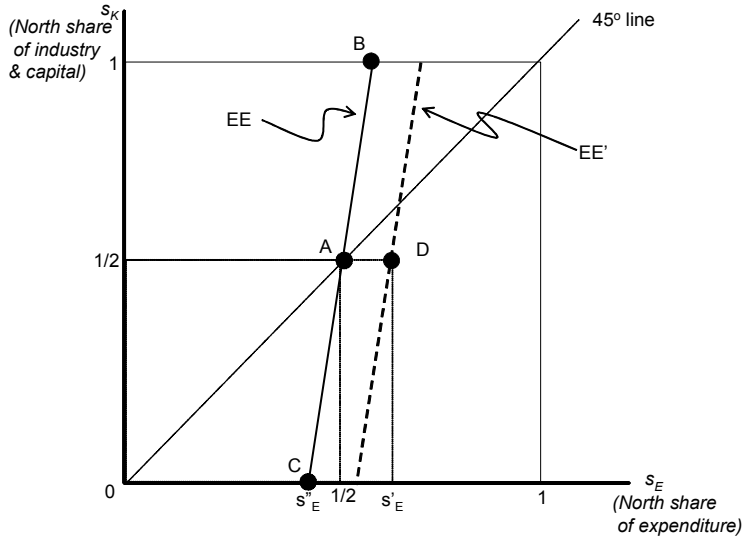
To focus on essentials, we work with only two regions, "north" and "south", and assume that they have the same technology and endowments. Furthermore, we assume that there are only two factors: labour, which is assumed immobile across regions, and capital, which is assumed very mobile across regions. In particular, capital flows to the region with the highest rate of return, so in equilibrium the rate of return is equalised across regions (or else all capital is in the high return region). There are two sectors, services and industry. Labour can work either in the service sector or in industry and we assume that industry is more capital intensive than services. Indeed, to minimise uninteresting complexity, we assume that each industrial firm requires some capital and some workers to produce its goods, while services are produced using labour alone. This means that a region's share of total capital is identical to the region's share of industrial firms.[1]

The logic of agglomeration and dispersion forces is best illustrated with a diagram that relates relative market size to the relative number of firms. The diagram has "s_E" (short for "share of expenditure") on the horizontal axis and this measures the relative market size of north. On the vertical axis is "s_K" (short for "share of capital") which shows the share of industry that is located in the north (the share of industry and the share of capital are identical as just mentioned).

[1] For a more detailed description of this type of analysis, see Chapter 3 in Baldwin et al. (2003).

We start out with perfect symmetry and rule out cost linkages by assuming that neither sector buys intermediate inputs. Consider first the demand linkage, i.e. the relationship between the share of industry in the north and the north's share of expenditure. Suppose industry—and thus capital—were evenly split between north and south. In this case, the two regions would have the same size markets as illustrated in Figure 2 by point A.

Figure 2. Demand linkages: The EE schedule



If all industry and thus all capital were in the north, s_E would be greater than ½, specifically point B in the diagram. In a similar fashion, point C shows the north's expenditure share when all the industry is in the south. The EE curve connecting these points shows the relationship between the north's share of industry and its share of capital when there are no inherent differences between north and south.

When south is fundamentally smaller than north, i.e. when in the initial situation, north has more than half the immobile factor, so the EE curve is shifted to the right as shown by EE'.

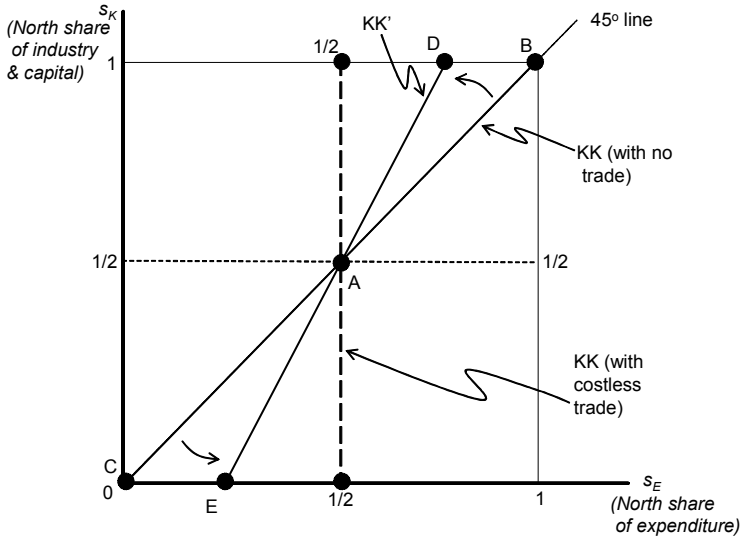
The second relationship between the two shares is the KK curve. Capital is mobile between regions and it moves to obtain the highest rate of return, so we need to calculate the rate of return in each region. The combinations of s_K and s_E that equalise rates of return is

called the KK curve. New economic geographers typically work with a simplified model where the reward to a unit of capital is proportional to firm-level sales.

It seems natural that equalising the profitability of the two regions would require the north's share of industry to rise as the north's share of expenditure rose. As argued above, firms that must choose one location will tend to prefer location in the big market, since this would allow them to economise on trade costs. But as more firms move into the big market, competition gets fiercer in the big market and gets weaker in the small market. Consequently, not all firms will move to the big market. The division of industry, i.e. s_K, adjusts to balance the agglomeration forces and dispersion forces. In the case where there is no trade between north and south, this will occur along the 45° line. With costly trade, however, it will occur along a line steeper than the 45° line because of the so-called home market effect.

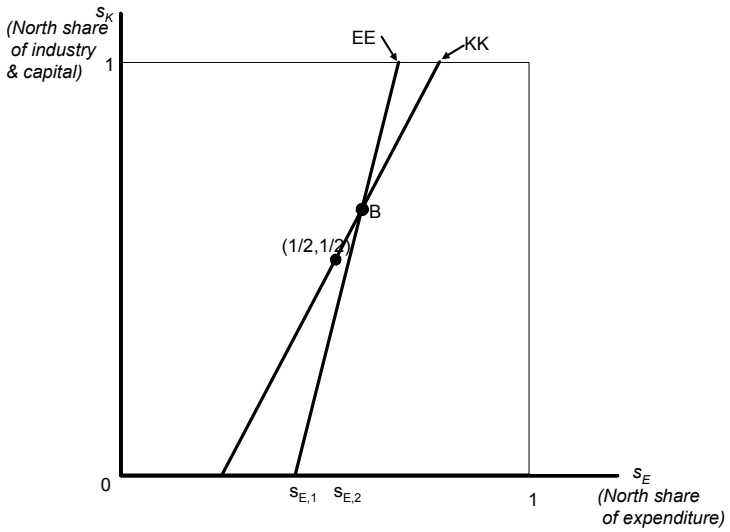
Starting at point A in Figure 3 and increasing north's expenditure share by 10 percent automatically reduces south's expenditure share by 10 percent. If s_K stayed at ½ when s_E was above ½ then the firms in the north would sell more than those in the south and thus earn more. To restore equal profitability, the degree of competition in the north would have to rise by 10 percent and the degree of competition in the south would have to fall by 10 percent. To do this, however, more than 10 percent of the firms need to move north. The point is that if the number of northern firms rises by 10 percent (by shifting firms from south to north), the degree of competition in the north will not rise by 10 percent. Why not? The reason is that northern firms now face lower competition in their export market—the southern market—since there are fewer locally based firms in the south. What this means is that restoring equal sales when there is trade will require the number of north-based firms to rise more than 10 percent. It is this bit of logic that is known as the "home market effect".

Figure 3. Equalised rates of return for capital: The KK schedule



Tighter economic integration rotates the KK line towards the vertical. The easiest way to see this is to contrast two extremes—the no-trade extreme, in which case the slope of the KK line is 45° as discussed above, and the costless trade case. When trade is costless, the division of firms between north and south is entirely irrelevant—any division would result in equal earnings per firm since each identical firm would sell the same amount in each region. Graphically, this is the vertical dashed line that extends from ½ to ½ as shown in the diagram. In other words, if the markets were of equal size (s_e =½), then any division of firm would equalise sales. A vertical KK line reflects this since any s_K works for a given s_E.

Figure 4. The locational equilibrium



Next we put together the EE and KK curves in Figure 4. The intersection of the EE and KK curves, point B, determines the equilibrium division of industry and the relative market sizes. In the case shown in Figure 4, north is fundamentally larger than south and will end up with the larger share of both industry and expenditure.

2.4. The impact of economic integration

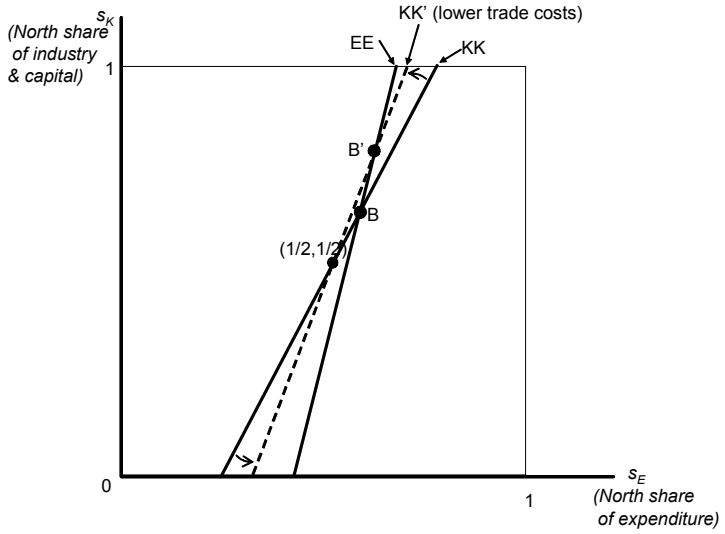
Finally, we are ready to consider the impact of deeper integration on the location of industry with the help of Figure 5. As trade costs fall, KK rotates counter-clockwise to KK' and the new equilibrium is B'. That is, tighter integration favours concentration of industry in the market that was initially bigger. Indeed, in this very simple model—where competition is the only anti-concentration force—continued lowering of trade costs leads to the "core-periphery" outcome. That is, a situation where all industry is in the big region (the core) and none is in the small region (the periphery).

Overall versus sectoral clustering

The logic that freer trade encourages agglomeration of industry in a particular region is quite robust. The outcome of this logic, however, can be very different depending upon how we interpret the resulting

clustering. As mentioned above, clustering takes two very different forms. Overall clustering leads to polarisation, i.e. big regional disparities in the levels of economic activity. But, the exact same logic can lead to sectoral clustering. That is, a circumstance where each region gets the "core" of one specific industry and becomes the "periphery" of other industries. The importance of this comment is that it suggests that increased specialisation by nation can be encouraged by agglomeration forces—not just comparative advantage forces.

Figure 5. Integration encourages geographic concentration



2.5. Adding back some elements of reality

In the EE-KK diagram, local competition is the only dispersion force so the model quite easily produces full agglomeration of capital/industry. In the real world, many things, especially land prices, tend to discourage full clustering. That is, as economic activity tends to cluster in, say, Paris, Parisian land prices rise and provincial land prices fall. This geographic change in the relative price of productive factors tends to prevent all activities from moving to the biggest market.

There are many other dispersion forces. For example, some types of industries are intensive in the use of natural resources that are immobile. Steel production, for example, tends to locate near iron ore

mines. Aluminum production, which requires huge inputs of electricity, tends locate near cheap sources of electricity, like hydroelectric dams and atomic energy plants.

3. First and a half nature

The two main elements in "1.5 nature geography" are the transportation network and factor endowments. Both of these are highly susceptible to government policy, but only in the fairly long run. The impact of the transport network is relative straightforward, so I concentrate on the factor endowments impact.

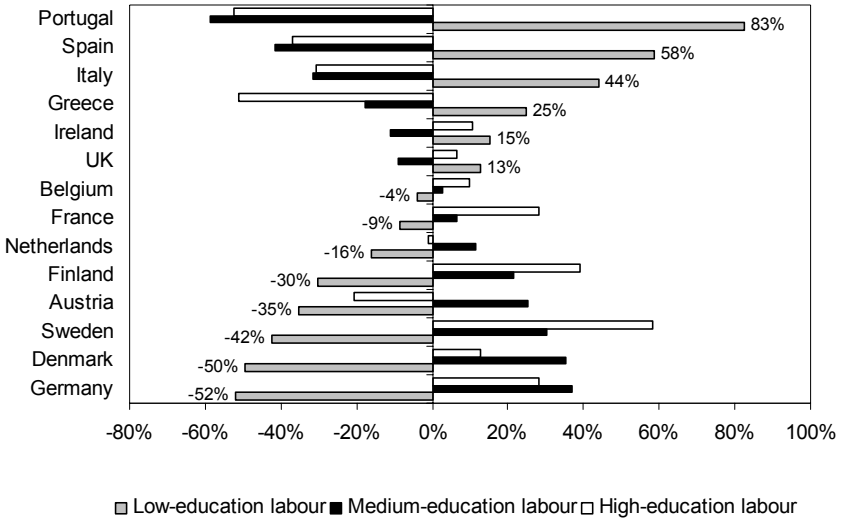
3.1. Comparative advantage

An elementary proposition in the theory of international trade is that liberalising trade raises economic efficiency by allowing each nation to concentrate its productive resources in sectors where they are relatively efficient, i.e. where it has a comparative advantage. This effect affects the location of industry; it tends to encourage sectoral specialisation nation-by-nation.

To make the basic point, consider the relative distribution of three types of labour in the EU: workers with little education (less than secondary), workers with at least secondary education, and highly educated workers (researchers). To make numbers comparable across nations of very different sizes, we compute each nation's supply of low-education workers relative to its total supply of workers and compare this to the same ratio calculated for the EU as a whole. The numbers are shown in Figure 6. For example, we see that Portugal's supply of low education workers (divided by Portugal total supply of workers) is 83 percent above the EU average. Germany's is 52 percent below the EU average.

Now consider what this means for the price of a good that uses low-education labour intensively, such as clothing. Without trade, Germany and Portugal would have to make all their own clothes. Since the factor that is used intensively in clothes production is relatively abundant in Portugal and relatively scarce in Germany, we should expect clothing to be more expensive in Germany than in Portugal, if there were no trade.

Figure 6. Relative labour endowments in Europe



Source: Data from Midelfart-Knarvik and Overman (2002); Data is average of 1993 and 1995.

When Germany and Portugal trade, Portugal exports clothing to Germany in exchange for goods that are relatively abundant in high-education labour. Using the same logic that told us clothing would be relatively cheap in Portugal without trade, we know that goods that are intensive in their use of high-education labour—for example, pharmaceuticals—would be relatively cheap in Germany. In this highly simplified world with trade only between Portugal and Germany, we would see Portugal exporting clothing (and other goods that are intensive in the use of low-educated labour) in exchange for pharmaceuticals (and other goods that are intensive in the use of high-educated labour) from Germany.

The spatial implications

In the example, trade induces an expansion in Portuguese sectors that are intensive in the use of low-education labour. Since the resources needed to expand output in these sectors must come from somewhere, trade also induces a contraction of other Portuguese sectors, in

particular, the sectors that had relatively high prices without trade, e.g. pharmaceuticals and other goods that are intensive in the use of high-education labour. In the simple example, the mirror-image shift would occur in Germany's industrial structure. If we view this from the international level, the resulting structural changes would look like a shift of clothing production from Germany to Portugal and a shift of the production of pharmaceuticals in the opposite direction. As a result, the industrial structures of both Portugal and Germany would become more specialised and industrial "clusters" would appear (clothing in Portugal and pharmaceuticals in Germany).

More generally, economic resources get shifted between sectors within each nation and, as a result, it looks like production is being reallocated sector-by-sector across nations. From the point of view of economic geography, this shows up as an increase in national specialisation sector-by-sector.

3.2. Transport networks

Market size is never an absolute in the modern world since trade is possible. The market available, for example, to firms located North of Stockholm is wider for goods that can be shipped by air than those that cannot (Stockholm's airport is north of the city). Since transport networks—especially road, rail and shipping networks—take a long time to establish and typically involve enormous sunk costs, the presence and nature of these networks acts very much like a skilled labour endowment. It is not immutable and it does affect the location of industry, but it is hard to change. Moreover, networks tends to get improved where there is a lot of economic activity and economic activity tends to be attracted to areas with good transport networks.

4. Concluding remarks

In the real world, the location of industry is determined by a tangle of factors the defies enumeration—everything from tax rates to beaches. It is therefore useful to abstract from reality and focus on the main forces. In this paper I suggest it is useful to organise the various causes into three main categories. The first concerns physical geography—like Stockholm's natural harbour; this is usually called first-nature geography. The second is the balance of economic agglomeration forces and dispersion forces. Most of these causes can be manipulated by policies such as production subsidies, trade liberalisation,

and taxation. These are usually called second-nature geography. I also suggest that there is an important in-between category, namely causes that adjust more slowly than industrial clusters but faster than coastlines. In the 1.5 geography, I would include transport networks and factor endowments. Both of these are subject to government policy—educational policies in the latter case and transport infrastructure policies in the former. It is intrinsically difficult to measure these empirically since they interact in complex ways that make it difficult if not impossible to determine what caused what in any particular region.

As is so often the case in economics, the things we care about the most are the hardest to sort out empirically. This leaves economists in the awkward position of either pretending that they know more than they do and making clear and compelling arguments for particular policies, or mumbling about "on the one hand … on the other hand." I hope that the rapidly growing empirical literature on economic geography will remove many of these doubts.

References

Baldwin, R. and Wyplosz, C. (2003), The Economics of European Integration, McGraw Hill, London.

Baldwin, R., Forslid, R., Martin, P., Ottaviano, G. and Robert-Nicoud, F. (2003), Economic Geography and Public Policy, Princeton University Press, Princeton.

European Commission (2003), Communication from the Commission: Second progress report on economic and social cohesion, COM (2003) 34 final.

Overman, H.G. and Midelfart-Knarvik, K.H. (2002), Delocation and European integration: Is structural spending justified?, Economic Policy 17, 321-359.

Starr, K., Balmford, M., Canterford, S. and Williams, C. (2004), Spatial patterns: A review of relevant theories, Mimeo, Bureau of Transport and Regional Economics,. Canberra.