6/12/2010

Relocating the value chain: off-shoring and agglomeration in the global economy*

Richard Baldwin,
Graduate Institute, Geneva and CEPR

Anthony J. Venables,
University of Oxford and CEPR

Abstract

Fragmentation of stages of the production process is determined by international cost differences and by the benefits of co-location of related stages. The interaction between these forces depends on the technological relationships between these stages. This paper looks at both cost minimising and equilibrium fragmentation under different technological configurations. Reductions in trade costs beyond a threshold can result in discontinuous changes in location, with relocation of a wide range of production stages. There can be overshooting (off-shoring that is reversed as costs fall further) and equilibrium may involve less off-shoring than is efficient.

Keywords: fragmentation, off-shoring, parts and components, assembly, vertical linkages, agglomeration.
JEL classification: F1, F23, R3

Authors' Addresses:

R. Baldwin	A.J. Venables
Graduate Institute	Department of Economics
Geneva	Manor Road
Switzerland	Oxford
	OX1 3UQ
	UK
Richard.Baldwin@graduateinstitute.ch	tony.venables@economics.ox.ac.uk

* Thanks to participants in seminars in Oxford and Princeton and conferences in Villars (CEPR), Rome (European Research Workshop in International Trade), Osaka (Asia Pacific Trade Seminar), and Jonkoping (European Regional Science Association).

1. Introduction:

An ever finer division of labour has been associated with economic progress for centuries. Its pace has varied; some authors speak of the industrial revolution as the first spatial unbundling (factories unbundled from consumers) with current phase of globalisation as the second unbundling (production stages unbundled across nations). The first unbundling was triggered by the revolutionary advances in transportation that came with steam power and subsequent advances in transportation technology. Transportation of goods and people is important in the second unbundling, but the trigger was the revolutionary advances in information and communications technology (ICT) that massively lowered the cost of organising complex activities over distances. The number of internet hosts took off from the mid-1980s and telephone penetration, which had been growing steadily, accelerated in the mid to late 1990s.

A consequence of cheap and reliable telecommunications combined with affordable high-capacity computing power was improved information management that transformed the organisation of group-work across space. Stages of production that had to be performed in close proximity – within walking distance to facilitate face-to-face coordination of innumerable small glitches – can now be dispersed without an enormous drop in efficiency or timeliness. Working methods and product designs have been shifted in reaction to the spatial separation, typically in ways that make production more modular. More recently, the second unbundling has spread from factories to offices, the result being the outsourcing and off-shoring of service-sector jobs.

Many of the international supply chains that developed are regional, not global. The cost and unpredictable delays involved in intercontinental shipping and travel of technicians and managers still matter. The first large-scale unbundling started in the mid 1980s and took place over short distances; the Maquiladora programme created 'twin plants', one on the US side of the border and one on the Mexican side. Although the programme existed since 1965, it only boomed in the 1980s with employment growing at 20% annually from 1982-89 (Dallas Fed 2002, Feenstra and Hanson 1996). Another unbundling started in East Asia at about the same time (and for the same reasons). In this region distances are short compared to the vast wage differences (Tokyo and Beijing are about 90 minutes apart by plane, yet in the 1980s the average Japanese income was 40 times the Chinese average). In Europe, the unbundling was stimulated first by the EU accession of Spain and Portugal in 1986, and then by the emergence of Central and Eastern European nations from the early 1990s.

Numerous examples serve to illustrate the pervasiveness of unbundling. The "Swedish" Volvo S40 has an air-conditioner made in France, the headrest and seat warmer made in Norway, the fuel and brake lines in England, the hood latch cable in Germany, and so on. Some parts are even made in Sweden (airbag and seat beats). These 'parts' are themselves made up of further parts and components, whose production is likely to be equally dispersed. For example, the air conditioner has a compressor, motor and a control centre, each of which may be made in a different nation. Sometimes these are owned or controlled by the original manufacturer, but often they are owned by independent suppliers.

Unbundling has been centre stage in much recent international trade research. There have been important empirical studies charting the rise of trade in parts and components (Ng and Yeats 1999, Hummels, Ishii and Yi 1999, Ando and Kimura 2005, Kimura, Takahashi and Hayakawa 2007). However, formal measurement has been problematic since trade data does not make clear what goods are input to other goods, and analyses based on standard techniques – input-output tables – are at too high a level of aggregation to capture the level of detail suggested by industry examples (Johnson and Noguera 2009).

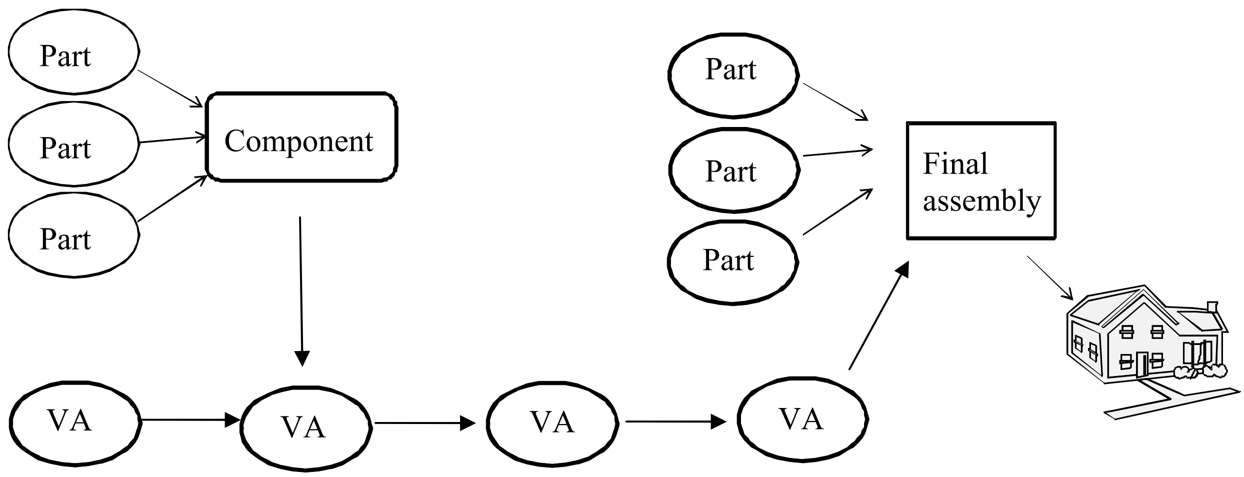
Analytical work has taken a variety of approaches. Much of the focus has been on taking simple characterisations of the technology of unbundling and drawing out the general equilibrium implications for trade and particularly for wages (Yi 2003, Grossman and Rossi-Hansburg 2008, Markusen and Venables 2007, Baldwin and Robert-Nicoud 2010). Others have linked it multinational activity (Helpman 1984, Fujita and Thisse 2006) and have placed it in the wider context of the organisation of firms (Helpman 2006).

This paper focuses on quite different aspects. We take seriously the fact that technology – the engineering of the product – dictates the way in which different stages of production fit together. Possibilities are illustrated in Figure 1. Each cell is a stage at which value is added to a good that ends up as final consumption, and each arrow is a physical movement of a part, component, or the good itself. They may be movements within a country (or within a plant), or may be unbundled movements between plants in different countries. There are two quite different configurations. One we refer to as the spider: multiple limbs (parts) coming together to form a body (assembly), which may be a component or the final product itself. The other is the snake: the good moving in a linear manner from upstream to downstream with value added at each stage.[1] Most production processes are complex mixtures of the two. Cotton to yarn to fabric to shirts is a snake like process, but adding the

[1] Dixit and Grossman (1982) undertake a general equilibrium analysis of the snake.

buttons is a spider. Silicon to chips to computers is snake like, but much of value added in producing a computer is spider-like final assembly of parts from different sources.

Figure 1: Snakes and spiders



In production processes like those illustrated in the diagram the location of any one element depends on the location of others. We suppose that costs are incurred if an arrow on the figure crosses an international boundary, and we will refer to these as 'off-shoring costs'. They are likely to be made up of costs of coordination and management as well as direct shipping costs. These off-shoring costs create centripetal forces binding related stages together.[2] Firms seek to be close to other firms with which they transact, but the form of this depends on the engineering of the product; it is different for snakes (linked to an upstream and a downstream stage) than for spiders. But there are also centrifugal forces that encourage dispersed production of different stages; for example, different stages have different factor intensities which create international cost differences and incentives to disperse. There is a tension between comparative costs creating the incentive to unbundle, and co-location or agglomeration forces binding parts of the process together.

Our objective is to analyse the interaction of these forces and show how they determine the location of different parts of a value chain. We look at the efficient location of these stages when decisions are taken by a single cost-minimising agent, and also at outcomes when stages are controlled by independent decision takers. Co-location and

[2] And costs associated with length and variability of time in transit, Hummels (2001), Harrigan and Venables (2006).

agglomeration forces mean that equilibrium is not necessarily cost minimising as coordination failures obstruct moves towards efficiency. There will be multiple equilibria and locational hysteresis. We show that the form this takes depends on engineering detail – snakes or spiders. This moves the paper significantly beyond earlier investigations of these issues which have often worked with highly symmetric and stylised (e.g. Dixit-Stiglitz) structures.[3]

The remainder of the paper develops models of the spider and the snake, looking at each in turn. While they give rise to different outcomes there are a number of general implications that we draw out in concluding comments. Throughout, the setting is a world of two countries, N and S and we will assume that all demand for the final product is in N.[4] Each stage of production – the cells of Figure 1 – can take place in either N or S, production costs differ between countries, and off-shoring costs are incurred in moving between N and S. The question is: what stages are produced in N, and what are produced in S?

2. The spider:

2.1: General setting

The spider is a production process in which parts are produced separately and come together for assembly.[5] Three sorts of cost may be incurred in getting a unit of final assembled product to the consumer; the cost of producing parts, off-shoring costs (i.e. costs of shipping and of coordinating off-shore production), and assembly costs.

Parts are indexed by type $y \in Y$, and the unit production cost of an individual part produced in S is $b(y)$. If the part is produced in N its production cost is normalised to unity (for all y). Minimum and maximum values of $b(y)$ are $\underline{b}$ and $\overline{b}$ and we assume that $\underline{b} < 1 < \overline{b}$, so low b parts can be produced more cheaply in S, and high b parts more cheaply in N. We will refer to low b parts as 'labour-intensive' (and high b parts as capital-intensive) although we are not explicit about the extent to which international cost differences are due to productivity or factor price differences. Parts come together for assembly, which requires

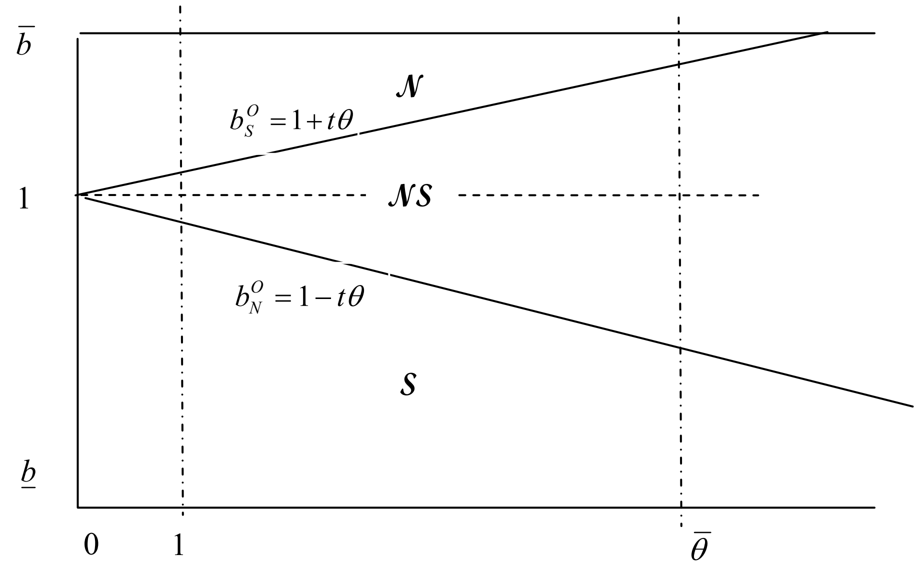
3 For example Venables (1996) and Fujita et al (1999).
4 It would be straightforward to a proportion of final demand in N and the remainder in S, but this assumption un-clutters the analysis.
5 We will think of this as assembly of the final product although it could be assembly of a component that is sold in N.

$\psi(y)$ units of each part $y \varepsilon Y$. Assembly also uses primary factors, and the units costs of these are a_N, a_S, according as assembly is in N or in S.

When a part's production is spatially separated from assembly a per-unit off-shoring costs of $t\theta(y)$ is incurred, this representing shipping costs and a wider set of communication, coordination and trade costs. The cost is the product of a parameter t capturing the overall level of off-shoring costs, and a type specific element $\theta(y)$, the support of which is $[1, \bar{\theta}]$. So, for example, high θ parts can be thought of as 'heavy' to ship and low θ parts as being 'light'. A final element of costs is incurred if assembly takes place in S; since the final market is in N a further off-shoring cost of $t.\alpha$ is paid. This is the product of parameter t and constant $\alpha > 0$ measuring the cost of off-shoring the final product; if $\alpha < \int_Y \theta(y)\psi(y)dy$ then it is cheaper to ship the assembled product than to ship all parts separately.

The combination of cost differences and off-shoring costs is illustrated on Figure 2 which has relative costs, b, on the vertical axis and the part specific element of off-shoring costs, θ, on the horizontal. The set Y is the area $[1, \bar{\theta}]$ x $[\underline{b}, \bar{b}]$ and each part can be represented as a point $\{b(y), \theta(y)\}$ in the set. The lines $b = 1 + t\theta$ and $b = 1 - t\theta$ divide the space into three sets according to the costs incurred in producing a part and delivering it to assembly. Parts in the upper region, $\mathcal{N}$, are more cheaply supplied from N, regardless of whether assembly is in N or S; even if assembly is in S, the cost of production in N plus off-shoring cost in reaching the assembler, $1 + t\theta$, is less than the cost of production in S, which is b. In the bottom region, $\mathcal{S}$, the converse is true: these are parts which are relatively labour-intensive and light, so are most cheaply supplied from S regardless of where assembly is; even if assembly is in N so off-shoring cost are incurred, supply from S is cheaper than supply from N, since $b + t\theta < 1$. The two lines dividing these sets can be thought of as 'off-shoring thresholds'. Values of b on the upper line will be labelled b_S^O, $b_S^O = 1 + t\theta$, since it gives the dividing line on which parts from N and S are equally expensive for delivery to assembly in S. The lower line $b_N^O = 1 - t\theta$, gives parts supplied at equal cost from N and S to assembly in N. Parts in the set $\mathcal{NS}$, between the lines, are most cheaply supplied if produced in the same place as assembly.

Figure 2: Production and off-shoring costs of parts



With this as set up, we now determine the location of assembly and associated location of parts. We do this in fairly general terms, and then in sections 2.2 and 2.3 move to particular distributions of parts over the space Y in order to get more specific results.

Single agent cost minimisation:

We look first at the case where a single agent (the assembler) determines the location of assembly and all parts in order to minimise total costs. When will the assembler choose to locate in N or in S? The answer is given by comparison of total costs in each situation, knowing that the assembler will locate production of each part to achieve the lowest delivered costs to the assembly plant. The two sides of the inequality below give total costs (assembly, parts, and off-shoring cost) when assembly is in N (left-hand side) and in S (right-hand side); it is cost minimising to assemble in S if the inequality is satisfied.

$$a_N + \int_{y \in \mathcal{N} \cup \mathcal{NS}} \psi(y)dy + \int_{y \in S} [b(y) + t\theta(y)]\psi(y)dy$$
$$> \quad a_S + \alpha t + \int_{y \in \mathcal{N}} [1 + t\theta(y)]\psi(y)dy + \int_{y \in S \cup \mathcal{NS}} b(y)\psi(y)dy \,. \qquad (1)$$

The left hand side of the inequality is assembly costs in N (equal to a_N) plus the costs of parts; types in sets ***N*** and ***NS*** are produced in N at unit cost 1, whilst those in set ***S*** are produced in S at unit cost $b(y)$ and also incur off-shoring costs $t\theta(y)$ to reach the assembler. The second line has the assembler operating in S at primary factor cost a_S with parts in sets ***NS*** and ***S*** produced in S and those in set ***N*** produced in N. Off-shoring costs are incurred on parts in set ***N*** and also on the assembled product that has to be shipped to the market in N.

Inequality (1) can be rearranged as

$$a_N - a_S + \int_{y\in NS} [1-b(y)]\psi(y)dy \;>\; t\left[\alpha + \int_{y\in N} \theta(y)\psi(y)dy - \int_{y\in S} \theta(y)\psi(y)dy\right]. \qquad (2)$$

The left hand side is the difference in production costs when assembly is in N rather than S, consisting of differences in assembly costs and in the production costs of parts in set ***NS*** which co-locate with assembly. Terms on the right hand side give the difference in off-shoring costs if assembly is in N compared to S; assembly in S incurs $t\alpha$ plus the costs of off-shoring parts from set ***N***, while saving off-shoring costs on parts from set ***S***.

By inspection of Figure 2 and equation (2) we see that, as $t \to 0$, all activities locate where their primary factor costs are lowest. The set ***NS*** disappears, so parts locate according to the value of $b(y)$ relative to unity, and assembly locates in S if $a_S < a_N$, and in N if $a_S > a_N$. At the other extreme, as $t \to \infty$, assembly and production of all parts takes place in N, where final demand is. The sets ***N*** and ***S*** disappear as all parts co-locate with assembly, and costs of off-shoring the assembled product, $t.\alpha$, come to dominate inequality (2).

While production costs determine location as $t \to 0$ and off-shoring costs are decisive as $t \to \infty$, at intermediate values of t there is tension between these forces. Consider first the case in which assembly is relatively cheap in S, $a_S < a_N$. We know from the preceding paragraph that assembly is in N at high t and in S at low t. As t falls the set ***S*** is enlarged ($b_N^O = 1 - t\theta$ on Figure 2 becomes flatter) so there is steady migration of parts from N to S, in line with their comparative costs. At some point it becomes cost minimising to relocate assembly to S and as assembly relocates so too do parts in set ***NS***, just leaving parts in ***N*** to be produced in N. However, further reductions in t enlarge set ***N*** (flattening $b_S^O = 1 + t\theta$) so some parts move back from S to N; lower off-shoring costs weakens the benefit of co-location relative to comparative production costs.

In the case in which assembly is relatively cheap in N, $a_S > a_N$, we know that assembly occurs in N at both very high and very low off-shoring costs. It may stay in N at all intermediate values, but it is also possible that as t is reduced it moves to S and then moves back to N, relocating parts in set $\mathcal{NS}$ as it does so. This happens if a high proportion of parts are labour intensive so that set $\mathcal{NS}$ is large. It is then efficient to move assembly and these parts to S, even though $a_S > a_N$. Assembly moves against its comparative costs, because the most efficient way to access parts produced cheaply in S (with low $b(y)$) is by moving assembly not by shipping parts. As t falls further the benefits of locating assembly with parts diminishes, and assembly moves back to N in line with its comparative costs. We work this out explicitly in section 2.2 using a particular distribution of parts across set Y.

Nash equilibrium in location:

What difference does it make if individual parts producers and the assembler take independent location decisions? Cost savings from co-location mean that there are potential coordination problems and, to set out the simplest case, we look at the simultaneous move Nash equilibrium.

Each parts producer takes the location of the assembler (and all other parts producers) as given, and locates where the unit cost of supplying the assembler is lowest. This gives location of parts as described by sets $\mathcal{N}$, $\mathcal{NS}$ and $\mathcal{S}$ on Figure 2. We assume that parts producers supply the assembler at cost.[6] The assembler takes as given the location of parts producers and chooses to locate in N or S to minimise overall costs.

Suppose first that assembly is in N, and that parts in sets $\mathcal{N}$ and $\mathcal{NS}$ are produced in N while those in set $\mathcal{S}$ are produced in S. There is no incentive for any parts producer to move, as each is in the country with the least cost of supplying the assembler in N. If the assembler switches location from N to S the location of parts producers is taken as unchanged, so the move is unprofitable (and assembly in N is an equilibrium) if

$$a_N - a_S \quad < \quad t\left[\alpha + \int_{y \in N \cup NS} \theta(y)\psi(y)dy - \int_{y \in S} \theta(y)\psi(y)dy\right]. \tag{3}$$

The left hand side is the change in the assembler's own primary factor costs. On the right hand side, the assembler now has to pay off-shoring costs on the assembled product and on

[6] This can be rationalised by a contestability assumption ensuring that parts' producers make zero profits. We relax it in section 2.3.

parts in sets $\mathcal{N}$ and $\mathcal{NS}$, while saving those in set $\boldsymbol{S}$.

Conversely, suppose that assembly is in S. In this case parts in set $\mathcal{N}$ will be produced in N while those in sets and $\boldsymbol{S}$ and $\mathcal{NS}$ are produced in S. If the assembler switches location from S to N (given location of parts producer) the deviation is unprofitable if

$$a_N - a_S \quad > \quad t\left[\alpha + \int_{y \in N} \theta(y)\psi(y)dy - \int_{y \in S \cup NS} \theta(y)\psi(y)dy\right] \tag{4}$$

i.e. the primary cost change exceeds off-shoring cost savings on the final product and on parts in $\mathcal{N}$, net of off-shoring costs incurred on products in sets $\boldsymbol{S}$ and $\mathcal{NS}$.

Two points follow immediately from these inequalities. First, equilibria do not necessarily deliver global cost minimisation; inequalities (3) and (4) are different from (2). And second, there may be multiple equilibria; it is possible that, for some parameters, inequalities (3) and (4) are both satisfied. The reason is simply that the location of production of parts in set $\mathcal{NS}$ is now taken as given instead of being directly controlled by the assembler. We draw out the implications in more detail in the section 2.3.

2.2. The spider: cost minimisation.

More explicit results can be derived by restricting attention to the case in which all parts have the same off-shoring costs (or 'weight'). We therefore replace $t\theta(y)$ by t, the same for all parts, collapsing the set Y down to one dimension. Since parts now vary in a single dimension the index y can simply be replaced by parts' b values (i.e. $y \equiv b$). Off-shoring thresholds are again denoted b_S^O, b_N^O, so set $\mathcal{N}$ is parts with b values greater than b_S^O, set $\boldsymbol{S}$ parts with b less than b_N^O, and set $\mathcal{NS}$ is parts with $b \in \left[b_N^O, b_S^O\right]$.

Total costs when assembly is in N (respectively S) are $C(b_N^O : N)$, $C(b_S^O : S)$, given by

$$\begin{aligned} C(b_N^O : N) &= a_N + \int_{b_N^O}^{\bar{b}} \psi(b)db + \int_{\underline{b}}^{b_N^O} (b+t)\psi(b)db \\ C(b_S^O : S) &= a_S + \alpha t + \int_{b_S^O}^{\bar{b}} (1+t)\psi(b)db + \int_{\underline{b}}^{b_S^O} b\psi(b)db, \end{aligned} \tag{5}$$

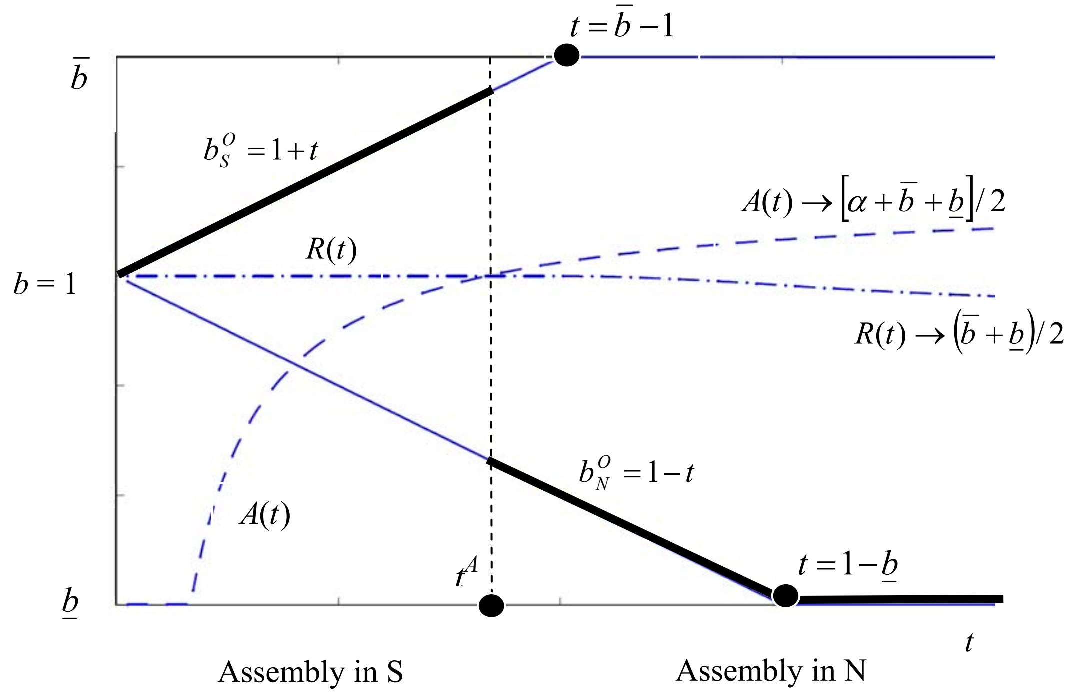
these equations corresponding to the two sides of inequality (1). The threshold values b_N^O, b_S^O

minimise the costs of supply in each case, and are therefore given by first order conditions

$$\partial C(b_N^O : N)/\partial b_N^O = \psi(b_N^O)\left[b_N^O + t - 1\right] = 0, \quad \partial C(b_S^O : S)/\partial b_S^O = \psi(b_S^O)\left[b_S^O - (1+t)\right] = 0. \qquad (6)$$

Of course, these correspond to the boundaries between regions on Figure 2. They are illustrated on Figure 3, analogous to Figure 2, but whereas Figure 2 was constructed for a given value of the off-shoring cost parameter *t* and a range $\theta(y)$, Figure 3 has $\theta(y) = 1$ and *t* varying on the horizontal axis. The off-shoring thresholds are functions of *t* given by the first order conditions (6), but constrained to lie in the support of *b*, so $b_S^O = \min[1+t, \bar{b}]$, $b_N^O = \max[1-t, \underline{b}]$. Thus, given *t*, parts on a vertical line in the interval $\left[\underline{b}, b_N^O\right]$ form the set ***S***; those in $\left[b_N^O, b_S^O\right]$ form set ***NS***; and those in $\left[b_S^O, \bar{b}\right]$ form set ***N***.

Figure 3: Cost minimising location, low cost assembly in S ($a_S < a_N$)



Under what circumstances will the assembler choose to locate in N or in S, given that she is controlling the location of all parts producers? The answer comes from comparison of total costs so, subtracting equations (5),

$$C(b_N^O : N) - C(b_S^O : S) = a_N - a_S + \int_{b_N^O}^{b_S^O} (1-b)\psi(b)db - t\left[\alpha + \int_{b_S^O}^{\bar{b}} \psi(b)db - \int_{\underline{b}}^{b_N^O} \psi(b)db\right] \quad (7)$$

If we further assume that each part in $[\underline{b}, \bar{b}]$ is used in the same quantity, $\psi(b) = 1$, this can be evaluated as

$$C(b_N^O : N) - C(b_S^O : S) = a_N - a_S + (b_S^O - b_N^O)\left(1 - \frac{b_S^O + b_N^O}{2}\right) - t\left[\alpha + \bar{b} + \underline{b} - b_S^O - b_N^O\right]. \quad (8)$$

This expression depends on *t*, and to analyse further it is convenient to split the right hand side into two parts, *R*(*t*) and *A*(*t*), constructed as

$$R(t) \equiv \left[b_S^O + b_N^O + \left(\frac{b_S^O - b_N^O}{t}\right)\left(1 - \frac{b_S^O + b_N^O}{2}\right)\right]\frac{1}{2} \quad (9)$$

$$A(t) \equiv \left[\frac{a_S - a_N}{t} + \alpha + \bar{b} + \underline{b}\right]\frac{1}{2}, \quad (10)$$

so $C(b_N^O : N) - C(b_S^O : S) = 2t[R(t) - A(t)]$. These two relationships are shown on Figure 3; it is cost minimising to locate assembly in S (equation (8) is positive) if *R*(*t*) lies above *A*(*t*) .

Notice that *R*(*t*) =1 if *t* is less than $\min[1-\underline{b}, \bar{b}-1]$ since in this range $b_S^O = 1 + t < \bar{b}$ and $b_N^O = 1 - t > \underline{b}$, and hence $b_S^O + b_N^O = 2$. Once *t* exceeds $\bar{b} - 1, 1 - \underline{b}$, b_S^O and b_N^O become respectively $\bar{b}$ and $\underline{b}$. *R*(*t*) then goes monotonically from unity to limiting value $(\bar{b} + \underline{b})/2$ as $t \to \infty$. The dependence of *A*(*t*) on *t* is clear from equation (10). As $t \to 0$, so *A*(*t*) tends to minus infinity if $a_S < a_N$, or to plus infinity if $a_S > a_N$. As $t \to \infty$, so *A*(*t*) goes monotonically to asymptotic value $[\alpha + \bar{b} + \underline{b}]/2$ which is, in all cases, greater than the asymptotic value of *R*(*t*). These observations are sufficient to give us information about potential intersections of the curves, as tabulated below:

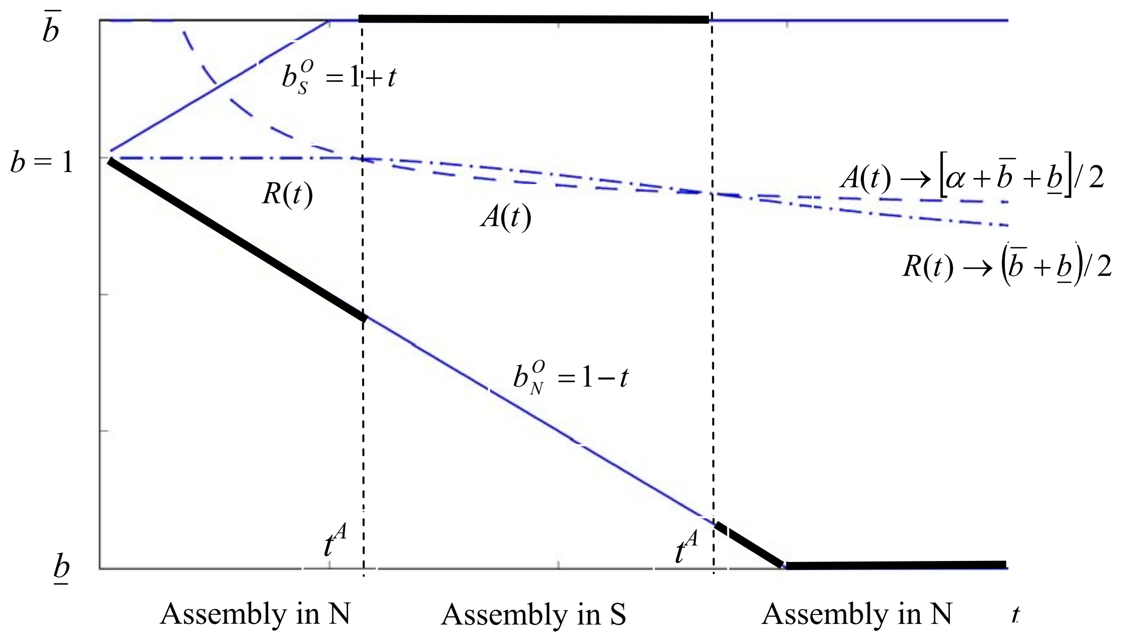
Table 1: Location of assembly

	$t \to 0$	$t \to \infty$	Intersections
$a_S < a_N$	$A(t) \to -\infty$ $< R(t) = 1$ Assembly in S	$A(t) \to \left[\alpha + \bar{b} + \underline{b}\right]/2$ $> R(t) \to \left(\bar{b} + \underline{b}\right)/2$ Assembly in N	1 intersection, t^A
$a_S > a_N$	$A(t) \to +\infty$ $> R(t) = 1$ Assembly in N	$A(t) \to \left[\alpha + \bar{b} + \underline{b}\right]/2$ $> R(t) \to \left(\bar{b} + \underline{b}\right)/2$ Assembly in N	0 or 2 intersections

For $a_S < a_N$, the case illustrated in Figure 3, there must be a single intersection of curves $R(t)$ and $A(t)$, as illustrated at t^A; to the right of this assembly is in N and to the left it is in S. The bold lines map out the location of parts as a function of t. Thus, if off-shoring costs fall through time location of the industry follows the pattern indicated by the these lines. Initially all production is in N, then declining t is associated with a slow migration of low b parts (labour intensive and below b_N^O) to S. At point t^A it becomes worthwhile to relocate assembly and a broad range of parts, $b_S^O - b_N^O$, from N to S. Some of these parts have higher costs in S than in N, but it is efficient to locate them close to the assembly plant in S; as t falls further, these parts move back from S to N as comparative factor costs becomes more important relative to off-shoring costs.

Figure 4 illustrates the case where N is the lower cost assembly location, $a_S > a_N$. High assembly costs in S mean that assembly takes place in N when t is very low, as well as when it is high; the function $A(t)$ is now decreasing in t, from plus infinity to its asymptotic value. If the cost advantage of S in parts is small, then $A(t)$ will lie above $R(t)$ everywhere and assembly will stay in N for all t. However, a large cost advantage for S means a lower value of the support of costs, $\left[\underline{b}, \bar{b}\right]$ and of $\underline{b} + \bar{b}$, shifting $A(t)$ down relative to $R(t)$ (see equations (9), (10)). There may then be two intersections of $A(t)$ and $R(t)$, as illustrated. It is then cost minimising, for an intermediate range of t, to move assembly and a substantial fraction (in Figure 4, all) parts production to S. The three phases can be summarised as follows: when t is high assembly stays close to the market because of costs of off-shoring the final product; at intermediate t it is cost minimising to use low cost parts producers in S and also to co-locate assembly in S with these parts producers; at low t all elements – parts and assembly – locate according to their comparative production costs. A necessary condition on parameters required for this double intersection is that $\left(\underline{b} + \bar{b}\right)/2 < 1$ so that, on average, parts are cheaper to produce in S than in N; a sufficient condition is given in the appendix, and is more likely to be met the lower are α and $a_S - a_N$ and the wider is the range $\bar{b} - \underline{b}$.

Figure 4: Cost minimising location, low cost assembly in N ($a_S > a_N$)



Production costs and off-shoring costs:

The preceding analysis (from the beginning of section 2.2) assumed that parts differ in their production costs, b, but all face the same off-shoring costs, t. What if there is a correlation between the two? This can be captured by making off-shoring costs a function of b, such as $t[1+h(b)]$.[7] We set $h(1) = 0$, so that if $h'(b) < 0$ then parts that are cheaper to produce in S have relatively high off-shoring costs, i.e. $h(b) > 0$ for parts with $b < 1$, and $h(b) < 0$ for $b > 1$. Our analysis will focus on this case, effects being reversed if $h'(b) > 0$. Equations (5) giving total costs when assembly is in N, S now become

$$C(b_N^O : N) = a_N + \int_{b_N^O}^{\bar{b}} \psi(b)db + \int_{\underline{b}}^{b_N^O} (b + t[1 + h(b)])\psi(b)db \tag{11}$$

$$C(b_S^O : S) = a_S + \alpha t + \int_{b_S^O}^{\bar{b}} (1 + t[1 + h(b)])\psi(b)db + \int_{\underline{b}}^{b_S^O} b\psi(b)db.$$

The off-shoring thresholds minimise the costs of supplying each location so, at an interior

[7] In terms of the general formulation of section 2.1 and Figure 2, the set Y is a line, but we no longer assume that it is vertical.

minimum, are

$$\partial C(b_N^O : N)/\partial b_N^O = \psi(b_N^O)\left[b_N^O - \left(1 - t[1 + h(b_N^O)]\right)\right] = 0 \tag{12}$$

$$\partial C(b_S^O : S)/\partial b_S^O = \psi(b_S^O)\left[b_S^O - \left(1 + t[1 + h(b_S^O)]\right)\right] = 0 .$$

The off-shoring thresholds therefore take value unity at $t = 0$, as in earlier cases. With $h'(b) < 0$ off-shoring thresholds are rotated downwards from this point. This is most easily seen by inverting the off-shoring thresholds to give $t = (1 - b_N^O)/[1 + h(b_N^O)] < (1 - b_N^O)$, $t = (b_S^O - 1)/[1 + h(b_S^O)] > (b_S^O - 1)$. This rotation is intuitive. It means that, if assembly is in N, then less is off-shored to S (set ***S*** is shrunk). Alternatively, if assembly is in S, more parts are produced in N (set ***N*** is enlarged).

What about the location of the assembler? Inspection of (11) indicates that having $h' < 0$, raises $C(b_N^O : N)$ and reduces $C(b_S^O : S)$ so increasing the incentive to off-shore. Proceeding as before, with $\psi(b) = 1$ and expressing the cost difference as $C(b_N^O : N) - C(b_S^O : S) = 2t[R(t) - A(t)]$, the function $R(t)$ is unchanged(equation 9) and $A(t)$ becomes

$$A(t) \equiv \left[\frac{a_S - a_N}{t} + \alpha + \bar{b} + \underline{b} + \int_{b_S^O}^{\bar{b}} h(b)db - \int_{\underline{b}}^{b_N^O} h(b)db\right]\frac{1}{2} . \tag{13}$$

If $h' < 0$ the first of the integrals is negative, and the second positive, so $A(t)$ is shifted downwards. From Figures 3 and 4, the implication is clear. In both cases there is an increase in the range of values of t at which assembly takes place in S. Contrary to what might be expected, relatively *higher* trade costs on labour intensive parts has the effect of *increasing* off-shoring to S. The intuition is that it is more expensive to access S's cost advantage just through off-shoring parts (and less expensive to access N's), and consequently more efficient to move assembly to S. Assembly moves to S, and so then do parts in the interval $[b_N^O, b_S^O]$. Conversely, if labour intensive parts have low off-shoring costs (h is increasing) then $A(t)$ is shifted upwards and it becomes more likely that assembly stays in N. Less off-shoring takes place, because it is efficient to keep assembly in N and just import the lowest cost parts from S.

Economies of scope in off-shoring:

A further possibility is that off-shoring costs depend on the range of parts that are produced in a particular location. For example, as more parts are off-shored so total off-shoring costs may increase less than proportionately due to economies of scope. These may arise in coordination and communication costs; coordinating the remote supply of two parts costs less than twice the cost of coordinating a single one. They may also arise because of costs associated with the possible disruption in supply of parts that are off-shored. Assembly might require that all parts are delivered on time, and is disrupted by a single part arriving late (Harrigan and Venables 2006). The probability of all parts arriving on time is decreasing but convex in the number of parts off-shored, so overall costs are increasing but concave.

To capture this we modify off-shoring costs by subtracting amount $t.g(.)$, where $g(.)$ is an increasing and convex function of the set of parts that are shipped to the assembler, with $g(0) = 0$. If assembly is in N and range $b_N^O - \underline{b}$ is produced in S, costs are reduced by $tg\left(b_N^O - \underline{b}\right)$, and similarly if assembly is in S. Total costs are then,

$$
\begin{aligned}
C(b_N^O : N) &= a_N + \int_{b_N^O}^{\bar{b}} \psi(b)db + \int_{\underline{b}}^{b_N^O} (b+t)\psi(b)db - tg\left(b_N^O - \underline{b}\right) \\
C(b_S^O : S) &= a_S + \alpha t + \int_{b_S^O}^{\bar{b}} (1+t)\psi(b)db + \int_{\underline{b}}^{b_S^O} b\psi(b)db - tg\left(\bar{b} - b_S^O\right).
\end{aligned} \tag{14}
$$

Off-shoring thresholds can be found as usual, and $g' > 0$ implies that they are steeper than they otherwise would be so that, given the location of the assembler, more parts are off-shored in order to get the marginal cost reduction from economies of scope. Setting $\psi(b) = 1$, and integrating, the additional terms can be put in $A(t)$ to give,

$$
A(t) \equiv \left[\frac{a_S - a_N}{t} + \alpha + \bar{b} + \underline{b} + g\left(\bar{b} - b_S^O\right) - g\left(b_N^O - \underline{b}\right)\right]\frac{1}{2} \tag{15}
$$

If unity is the midpoint of the support $[\underline{b}, \bar{b}]$ then Figures 3 and 4 are symmetric around $b = 1$, $g\left(\bar{b} - b_S^O\right) = g\left(b_N^O - \underline{b}\right)$ and economies of scope have no impact on the location of the assembler. Figures 3 and 4, as drawn, have a wider range of parts produced more cheaply in S than in N (unity greater than the midpoint of $[\underline{b}, \bar{b}]$), implying that at each t,

$g(\bar{b} - b_S^O) < g(b_N^O - \underline{b})$. $A(t)$ is then shifted down, this increasing the range of t values over which assembly takes place in S; there is more potential to reap economies of scope in off-shoring in S than there is in N.

2.3. The spider: equilibrium.

We now turn from overall cost minimisation to the Nash equilibrium in which each firm (assembler and each separate part) makes its location choice given the location of all others. Inequalities (3) and (4) give the conditions under which a deviation by the assembler is not profitable, and we now apply these to the case in which off-shoring costs are the same for all parts and $\psi = 1$. The equilibrium location of parts producers is given by the sets $\boldsymbol{\mathcal{N}}$, $\boldsymbol{\mathcal{N}S}$ and $\boldsymbol{S}$, defined by off-shoring thresholds b_S^O, b_N^O. Inequality (3), the condition for assembly to be in N, becomes

$$a_N - a_S \quad < \quad t\left[\alpha + \left(\bar{b} - b_N^O\right) - \left(b_N^O - \underline{b}\right)\right], \text{ or, using (10)}, \quad A(t) \quad > \quad b_N^O, \qquad (16)$$

while that for assembly in S, inequality (4) is

$$a_N - a_S \quad > \quad t\left[\alpha + \left(\bar{b} - b_S^O\right) - \left(b_S^O - \underline{b}\right)\right], \text{ or, using (10)}, \quad A(t) \quad < \quad b_S^O. \qquad (17)$$

Equilibria can therefore be illustrated by comparison of $A(t)$ with the off-shoring thresholds, and are shown on Figures 5 and 6, constructed with the same parameter values as Figures 3 and 4. The lower bold line on each figure gives equilibrium with assembly in N; inequality (16) is satisfied so it is cost minimising for the assembler to locate in N, and most parts (all those in interval $[b_N^O, \bar{b}]$) are produced in N. The upper bold line gives equilibria with assembly in S; most parts (those in interval $[\underline{b}, b_S^O]$) are produced in S and it is consequently cost minimising for assembly to locate in S (inequality (17) is satisfied).

Figure 5 deals with the case where assembly is cheaper in S than in N. It illustrates that at high or medium off-shoring costs there are two equilibria. Assembly in S implies that a high share of parts production is in S, this supporting the choice of the assembler to locate in S. Similarly, assembly in N is supported by the presence of many parts producers in N. Starting from a high value of t and with all production in N, reductions in t move the equilibrium along the lower heavy line until point Ω at which the equilibrium with assembly in N ceases to exist, and assembly jumps to S. The jump relocates a set of parts producers some of which move back to N as further reductions in t make co-location with the assembler less important. The small circle gives, for comparison, the jump point from Figure 3 (corresponding to point t^A) . It indicates that, during a process in which assembly starts in N and t is falling, equilibrium assembly moves to S later than is efficient.

Figure 5: Equilibrium locations, low cost assembly in S ($a_S < a_N$)

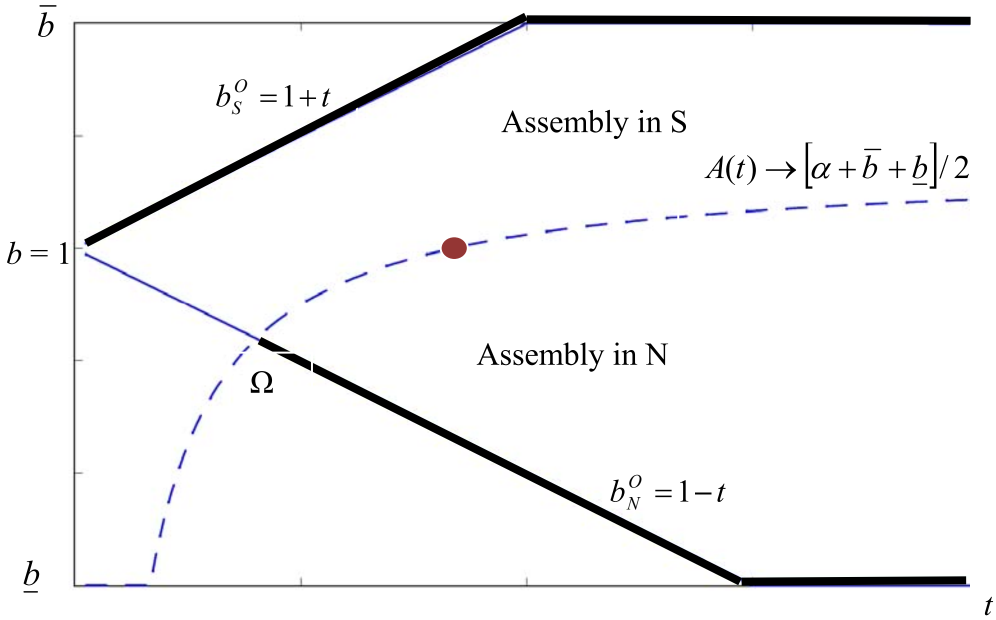
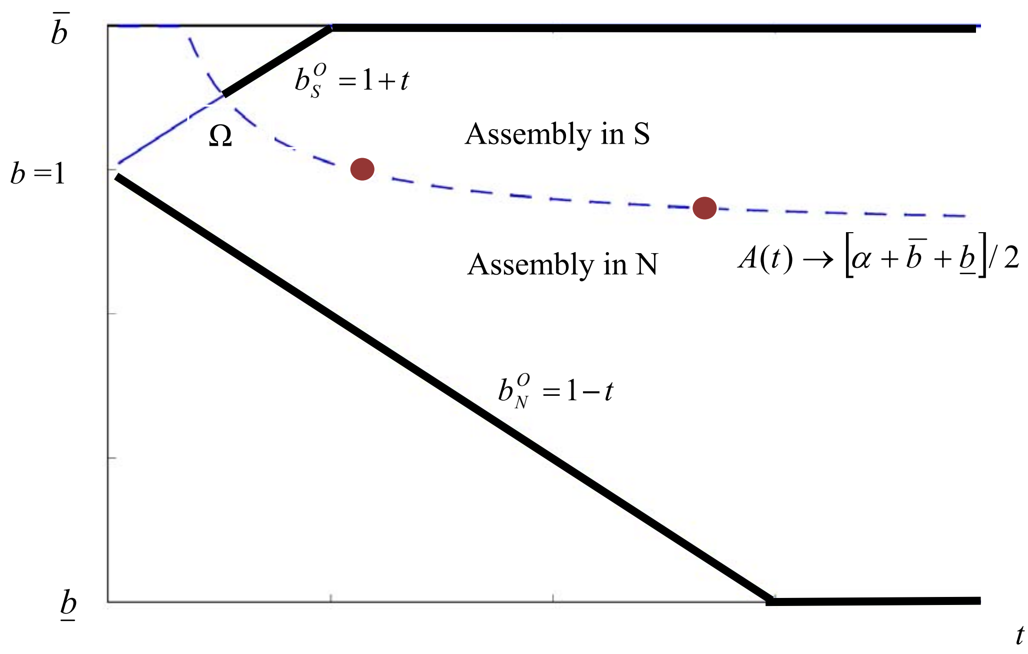


Figure 6 gives the case where primary factor costs of assembly are higher in N than in S. As illustrated assembly in N is an equilibrium at all values of t; starting from this position, off-shoring of parts takes place in a continuous manner but assembly never moves. However, there is, for t greater than Ω, an alternative equilibrium in which assembly in S is supported by location of a high share of parts production in S. Since Figure 8 has the same parameters as 4, so we know that it is efficient for assembly to locate in S in the interval between the circles. However, inertia due to co-location effects means that this does not occur. It is possible to make an example in which there is an interval of t in which the equilibrium with assembly in N does not exist, ($A(t)$ intersects $b_N^O = 1 - t$ twice), this requiring a parameter change that shifts $A(t)$ downwards (e.g. a lower value of α or lower support $[\underline{b}, \bar{b}]$).

Figure 6: Equilibrium locations, low cost assembly in N ($a_S > a_N$)



Pricing in equilibrium:

Analysis of the equilibrium has assumed that parts are sold to the assembler at unit cost, suggesting that this can be supported by contestability – the possibility that entry of parts producers can occur to bid away any profits. This assumption sits uncomfortably with Nash equilibrium in location in which the assembler takes the location of parts producers as given. The former assumes that entry is anticipated in response to price change, while the latter assumes entry is not anticipated in response to location change.

One response to this is to draw out the fact that price changes and location changes are quite different activities with different implications. It is possible to justify the configuration of assumptions by drawing on this difference. For example, contestability might occur only once a part producer is already present in a location; the presence of one producer of a particular part introduces the technology and makes copying and subsequent entry easy.

Another response is to model an endogenous price-cost mark up and test how this modifies outcomes. To show how this can be done we maintain contestability in N (where there are many firms and access to technology is easy) but suppose that if a part is produced in S then it is done so by a single firm. The firm faces no threat of local competition, but faces competition from potential supply in N. The price charged for a part produced in S is determined by a Nash bargain between the assembler and parts producer, where the surplus to be divided is the difference between cost of supply from N and from S. The magnitude of this surplus depends on where the assembler is located, and we denote the price charged by a parts producer in S when the assembler is in N (respectively S), p_N , (p_S), given by:

$$p_N = b + \gamma(1-(b+t)), \qquad p_S = b + \gamma(1+t-b). \tag{18}$$

In these expressions γ is the share of the parts producer in the bargain. If the assembler is in N the surplus from producing the part in S is $1 - (b + t)$; share γ is captured by the parts producer who therefore receives p_N, while the assembler pays $p_N + t$. If the assembler is in S the surplus is $1 + t - b$; share γ goes to the parts producer who receives p_S, equal to the amount paid by assembler since they are co-located.

Given this pricing, we seek to establish the equilibrium location of parts production and the associated levels of assembler's costs. Suppose first that the assembler is in N. Parts are produced in S if they make non-negative profits, i.e. $p_N - b = \gamma(1-(b+t)) \geq 0$. This implies that the off-shoring threshold is $b_N^O = 1-t$, exactly as before, reflecting the fact that there is no surplus at the margin (and implying the Nash bargain is efficient). Assembler's costs are

$$C(b_N^O : N) = a_N + \left(\overline{b} - b_N^O\right) + \int_{\underline{b}}^{b_N^O} (b + \gamma(1-(b+t)) + t)db \tag{19}$$

where terms on the right hand side are direct assembly costs, costs of parts supplied from N at price unity, and cost (to the assembler) of parts supplied from S. If the assembler is in S, parts production in S makes non-negative profits if $p_S - b = \gamma(1+t-b)) \geq 0$. The off-shoring threshold is $b_S^O = 1+t$ and assembler's costs are

$$C(b_S^O : S) = a_S + \alpha t + (1+t)\left(\overline{b} - b_S^O\right) + \int_{\underline{b}}^{b_S^O} (b + \gamma(1+t-b))db \tag{20}$$

We want to know when a deviation by the assembler from N to S reduces costs, given the location of parts producers, i.e. when $C(b_N^O : N) - C(b_N^O : S) > 0$. Subtracting (20) from (19), but with (20) evaluated at b_N^O not b_S^O gives

$$C(b_N^O : N) - C(b_N^O : S) = a_N - a_S - \alpha t - t\left(\bar{b} - b_N^O\right) + \int_{\underline{b}}^{b_N^O} t(1-2\gamma)db$$

$$= a_N - a_S - t\left[\alpha + \left(\bar{b} - b_N^O\right) - \left(b_N^O - \underline{b}\right)(1-2\gamma)\right] \qquad (21)$$

To compare this with the previous section, the condition for a deviation from N to S to be profitable is that *t* has fallen to the level at which

$$A(t) \;\; > \;\; (1-\gamma)b_N^O + \gamma \underline{b} \qquad (22)$$

If $\gamma = 0$ this reduces the previous case (equation (16)), as it should. Otherwise, assembly in N remains an equilibrium for a wider range of values of *t*, as is clear from inspection of Figure 5. The intuition is that off-shoring moves the assembler into a less competitive environment; higher γ therefore further postpones off-shoring, amplifying the inefficiency we saw in the previous section.

2.3. The spider: conclusions.

We have shown that a process of globalisation which brings reductions in *t* leads to quite rich patterns of off-shoring. If the industry is controlled by a single cost minimising agent then falling *t* over some intervals causes a steady migration of parts, but will then cross thresholds where assembly relocates, taking with it a wide range of parts. These movements may appear to be against the comparative costs of an element that is moving, because of co-location effects. Thus, a capital intensive part may move to S when assembly moves. Furthermore, it is possible that assembly moves to S to better access labour intensive parts, even if the primary factor costs of assembly are lower in N. A corollary is that there is ‘overshooting’; as *t* falls further co-location benefits are reduced and comparative production costs become the decisive factor determining location of each stage.

The Nash equilibrium may fails to achieve global cost minimisation. The reason is coordination failure. If the assembler takes the location of parts producers as fixed then the benefits of a coordinated move (assembler and a range of parts) cannot be achieved. If *t* is falling through time, off-shoring will occur inefficiently late.

3. The snake:

We now turn to the case where the product moves through a vertical production process with value being added at each stage. The stages form a continuum indexed z ε (0,1) with $z = 0$ the most upstream stage and $z = 1$ the final and most downstream. Every stage combines primary factors with the output of the previous stage. Techniques of production and off-shoring costs may vary across stages and, since the technology of production determines the ordering of stages, neither factor intensity nor off-shoring costs need vary continuously with z. Nevertheless, we think that considerable insight can be got by making these characteristics depend continuously (and in cases studied below, monotonically) on z. We will show how outcomes differ according to the relative production costs and off-shoring costs of upstream and downstream stages.

Primary factor costs incurred at stage z will be set equal to unity in N and denoted $c(z)$ in S, so a low c value denotes a stage with high labour intensity or productivity advantage in S. The full cost of the product at stage z is cumulative value added, the integral of primary costs over upstream stages. Thus, if all production is in N, full cost at stage z is equal to z.

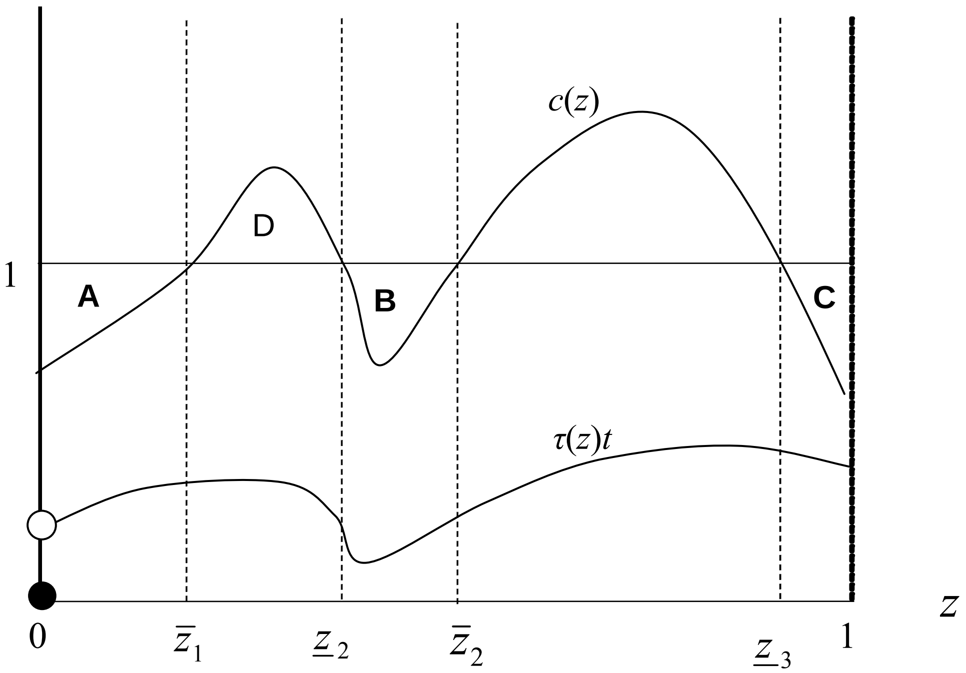
Off-shoring costs are incurred where production switches location, with per unit cost denoted $\tau(z)t$. As before, t is a parameter common to all stages and capturing the overall technology of off-shoring. Off-shoring costs may vary with z, and the function $\tau(z)$ can be thought of as capturing how the weight of the product varies along the production chain. It may be increasing as value is added, or decreasing (e.g. in refining or smelting a natural resource). An important criterion turns out to be whether 'weight' is added more or less fast than value added. All final consumption is in N, as before.

Figure 7 illustrates costs along the snake, and we start with informal discussion of this example before moving to formal analysis. The horizontal axis is stages of production, z, and the horizontal line is the unit cost (value added) of each stage if undertaken in N. Wiggly lines $c(z)$ and $\tau(z)t$ are unit costs in S and off-shoring costs respectively.[8] As illustrated, stages in the ranges marked A, B, C are 'labour intensive' with lower cost in S. Should they be undertaken in S, given that other segments are undertaken in N? If the interval A is off-shored the production cost saving is given by the area A, while off-shoring costs $\tau(\bar{z}_1)t$ are incurred. Notice that there may be a discontinuity in these costs, since $\tau(0) = 0$ (no off-

[8] Not monotonic, but drawn to be continuous.

sourcing costs are incurred if no stages are produced in S), while it is possible that $\lim_{z \to 0} \tau(z) > 0$ (moving even the smallest stage to S incurs off-shoring costs). Interval B is a range of production cost saving, but if undertaken in S while stages on either side of B remain in N, off-shoring costs at both ends of the interval are incurred, $[\tau(\underline{z}_2) + \tau(\bar{z}_2)]t$. As illustrated, it is certainly not cost minimising to move the whole interval B to S; shifting $\underline{z}_2$ slightly to the right has no impact on production cost savings (around $c(\underline{z}_2) = 1$), and brings a finite saving in off-shoring costs. Notice, also that, if $[\tau(\bar{z}_1) + \tau(\underline{z}_2)]t > D$ it may be efficient to move the whole range A + D + B to S, even if it is not efficient to move A and B separately. Finally, range C: the assumption that all final consumption is in N means that cost savings have to be weighed against off-shoring costs at both ends of the range, $\tau(\underline{z}_3)t$ and $\tau(1)t$. It can only be cost minimising to produce this range in S if $\underline{z}_3$ is a discrete distance below unity, so that the costs of off-shoring are offset by cost saving across a relatively wide range of stages.

Figure 7: Costs along the snake



With this as introduction, we now move to a more formal analysis comparing cases in which off-shoring commences from the upstream end of the snake, and where it starts downstream. These cases arise if cost differences are monotonic with, for example, S's cost advantage steadily increasing or decreasing over the stages of production.

3.1 The snake: cost minimisation:

What is the efficient location of stages between N and S, from the standpoint of a single cost-minimising agent? There are qualitatively different cases, depending on whether production in S is most valuable for upstream (case U) or for downstream stages (case D). We look first at case U, and suppose that $c(z)$ is increasing, with $c(0) < 1$ and $c(1) > 1$, so S has a production cost advantage in upstream stages and disadvantage in downstream stages. In case U the cost minimisation problem can be set up by supposing that upstream stages $z \leq \hat{z}$ take place in S, downstream stages $z > \hat{z}$ take place in N, and the problem is to choose $\hat{z}$ to minimise total costs, denoted $U(\hat{z})$, and given by

$$\hat{z} = 0: \quad U(\hat{z}) = \int_0^1 dz = 1$$

$$\hat{z} \in (0,1]: \quad U(\hat{z}) = \int_0^{\hat{z}} c(z)dz + \tau(\hat{z})t + \int_{\hat{z}}^1 dz . \tag{23}$$

If $\hat{z} = 0$ then all production takes place in N and total costs are unity. If $\hat{z} > 0$ the second equation holds, in which the first integral is the cost of producing the range $z \leq \hat{z}$ in S; $\tau(\hat{z})t$ is the cost of transferring the product to N, and the final integral is the sum of the (unit) cost of producing remaining stages in N.[9] We assume that the functions $c(z)$ and $\tau(z)$ are twice differentiable, so first and second derivatives with respect to $\hat{z}$ are

$$\frac{\partial U(\hat{z})}{\partial \hat{z}} = c(\hat{z}) + \tau'(\hat{z})t - 1 . \tag{24}$$

$$\frac{\partial^2 U(\hat{z})}{\partial \hat{z}^2} = c'(\hat{z}) + \tau''(\hat{z})t . \tag{25}$$

A number of issues arise in analysing the cost minimising $\hat{z}$ and its dependence on t.

First, under what conditions does the first order condition $\partial U(\hat{z}) / \partial \hat{z} = c(\hat{z}) + \tau'(\hat{z})t - 1 = 0$ give a local minimum? From (25), a sufficient condition is

[9] The objective has to be written in these two parts because of the discontinuity in $\tau(z)$ noted above.

that $c'(\hat{z}) > 0$ and $\tau''(\hat{z}) > 0$. The former says, that S has comparative advantage in upstream products, so it is these that our off-shored. The second rules out situations with sharply falling off-shoring costs, like $\underline{z}_2$ on Figure 7.

Second, if the first order condition defines a local cost minimum, how does $\hat{z}$ vary with the overall level of off-shoring costs, *t*? Differentiating along the first order condition,

$$\frac{d\hat{z}}{dt} = \frac{-\tau'(\hat{z})}{c'(\hat{z}) + \tau''(\hat{z})t}. \qquad (26)$$

The denominator is positive by the second order condition, but the numerator may be positive or negative. If $\tau'(z) > 0$, so the product is getting 'heavier', then a reduction in *t* increases the range of stages undertaken in S; however, if $\tau'(z) < 0$, as for a primary product that is losing weight, then reducing in *t* reduces the range of activities in S, since shipping costs are less of an obstacle to shifting to N.

Third, the local minimum may not be a global minimum, because of the discontinuity at zero, $\tau(0) = 0$ but $\lim_{z \to 0} \tau(z) > 0$. In this case values of *U* at zero and at the local minimum must be evaluated and compared. It will then be the case that when *t* is high all production will be in N (since any production in S incurs off-shoring cost proportional to *t*) , but when *t* is low comparative costs will give an interior cost minimum, with upstream stages in S and downstream in N.

Finally, the turning point will not be a local minimum if the second order condition fails. If we maintain the assumption that S has a comparative advantage in upstream goods, $c'(\hat{z}) > 0$, then this will occur if $\tau''(\hat{z}) < 0$ and *t* is large enough. We therefore expect the cost minimising outcome to be a corner solution when *t* is large, and an interior solution when it is small.

These points indicate a wide range of possibilities which we illustrate by supposing that *c*(*z*) is increasing linearly in *z*, and that $\tau(z) = \theta_0 + \theta_2 z + \theta_2 z^2$, looking at alternative values of coefficients. Figure 8 below gives four cases(coefficients are given in the appendix). The bottom panel gives the alternative shapes of $\tau(z)$ that we explore, with *z* on the horizontal and $\tau(z)$ on the vertical. The top panel gives the relationship between the cost-minimising location of stages and *t*; it has *t* on the horizontal axis and $\hat{z}$, the range produced in S, on the vertical.

The most straightforward case is the solid line labelled (1) in each figure. There is no discontinuity in trade costs, the second order condition holds and $\tau'(z) > 0$ throughout. Off-shoring is given by the first order condition and we see on the upper panel that, as *t* falls, a steadily increasing share of downstream stages are produced in S. The second case is the short-dashed line labelled (2). Any off-shoring incurs some off-shoring costs (which then increase the more stages are off-shored. The discontinuity means that it is never cost minimising to off-shore a very small number of stages, so off-shoring, when it commences, takes place at finite scale as indicated by the jump in the upper panel.

The third possibility is that that the product gets 'lighter' the more processing is done; we capture this by the long-dashed line, case (3); this case must have $\lim_{z \to 0} \tau(z) > 0$ to be consistent with $\tau'(z) < 0$. We therefore have discontinuity giving the jump, as in case (2), and comparative statics giving a smaller range produced in S the lower is *t*. At high *t* the off-shoring margin is increased to get the advantage of lower τ, a force that is weakened as *t* falls, so lower off-shoring costs reduce the amount of production in S.

Finally, case (4), in which off-shoring costs are (strongly) concave, with $\lim_{z \to 0} \tau(z) > 0$ and $\tau''(z) < 0$. At high *t* everything is produced in N; when off-shoring costs become low enough all stages of production occur in S, since $\tau(1)$ is small relative to intermediate values of τ (the first order condition is a local cost maximum). Lower values of τ change the sign of the second order condition (25) from negative to positive, so the first order condition comes to define a cost minimum. As in the preceding section, there is 'overshooting' as falling trade costs make off-shoring possible, but co-location forces mean that it is not yet fully aligned with the comparative costs of each stage.

Figure 8a: Upstream off-shoring

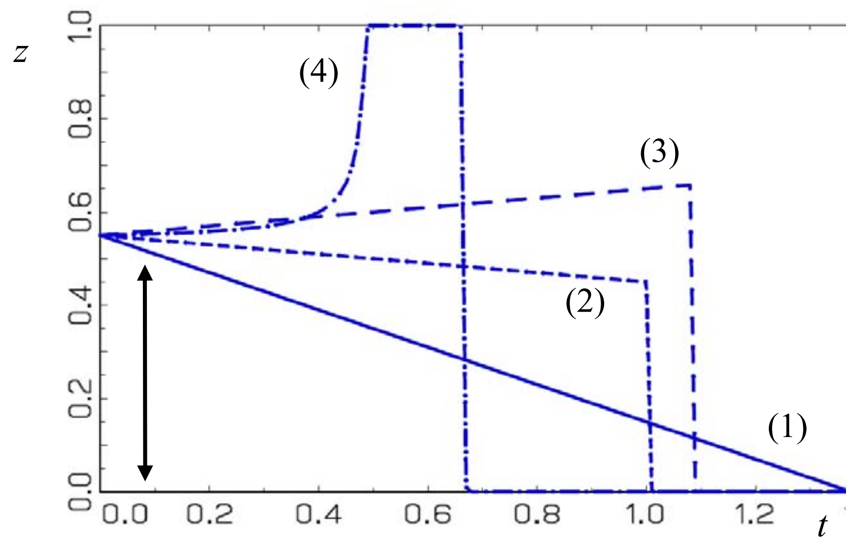
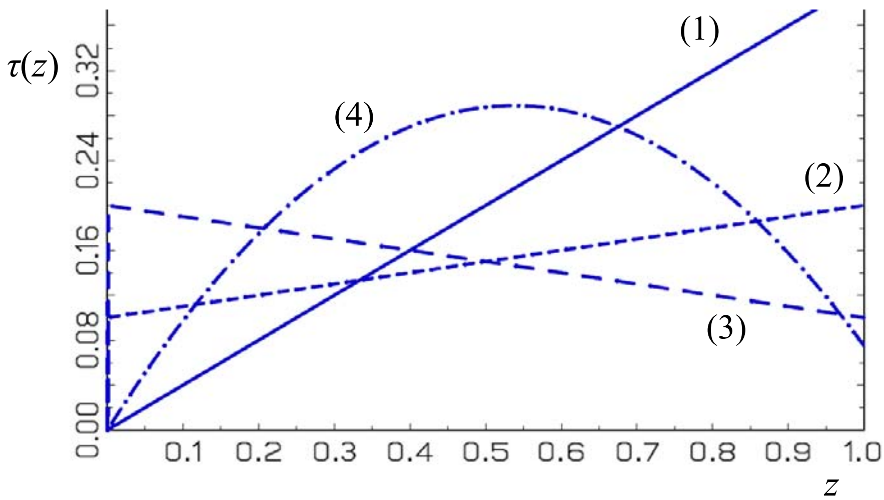


Figure 8b: Functions τ(z)



This analysis assumed that off-shoring occurs for an upstream set of stages, validating this with S having a cost advantage in upstream stages, $c'(z) > 0$, $c(0) < 1$ and $c(1) > 1$. What about the converse case, where off-shoring, if it occurs, takes place for downstream stages? We now reverse comparative costs and suppose that $c'(z) < 0$, $c(0) > 1$ and $c(1) < 1$. The cost minimisation problem is that of minimizing total costs $D(\tilde{z})$ where (opposite to case U) stages upstream of $\tilde{z}$ ($z \leq \tilde{z}$) takes place in N and downstream of $\tilde{z}$ ($z > \tilde{z}$) take place in S. $D(\tilde{z})$ takes the form:

$$\tilde{z} = 1: \quad D(\tilde{z}) = \int_0^1 dz = 1 \qquad (27)$$

$$\tilde{z} \in (0,1]: D(\tilde{z}) = \int_0^{\tilde{z}} dz + \tau(\tilde{z})t + \int_{\tilde{z}}^1 c(z)dz + \tau(1)t .$$

Notice that off-shoring a downstream range of parts (an interval [*z*, 1], *z* > 0) has an important difference from the previous case because (since consumption only takes place in N) it incurs *double* transport costs, τ(z)*t* plus τ(1)*t*. As before, there is a discontinuity where off-shoring commences, with $D(\tilde{z})$ jumps upward as $\tilde{z}$ goes below unity. Its gradient is then, for $\tilde{z} \in (0,1]$,

$$\frac{\partial D(\tilde{z})}{\partial \tilde{z}} = 1 - c(\tilde{z}) + \tau'(\tilde{z})t . \qquad (28)$$

and convexity requires

$$\frac{\partial^2 D(\tilde{z})}{\partial \tilde{z}^2} = -c'(\tilde{z}) + \tau''(\tilde{z})t > 0 . \qquad (29)$$

Figure 9: Downstream off-shoring

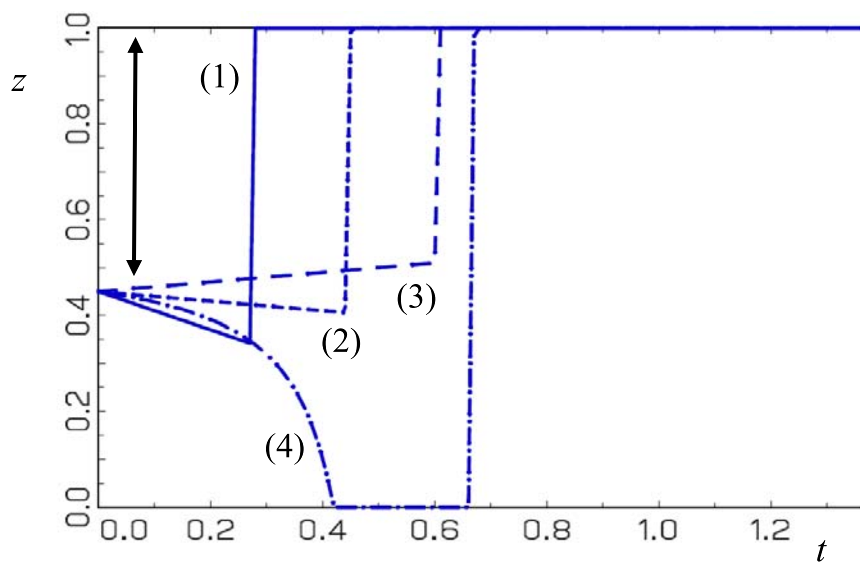


Figure 9 illustrates outcomes with *c*(*z*) linearly decreasing, and τ(*z*) taking the same four shapes as given in Figure 8b. Three points stand out. First, if τ(1)*t* > 0, off-shoring is necessarily discontinuous. A finite range of cost saving stages has to be off-shored to cover the double transport costs incurred. Second, once downstream stages are off-shored the effect of further reductions depends on $\tau''(\hat{z})$; in case (1) products are getting heavier as they move down the snake, so creating a marginal incentive to do more off-shoring; a reduction in *t* reduces this incentive so causes less outsourcing. Third, the second order condition depends on *t* and may fail, as in case (4); at intermediate *t* it is profitable to off-shore all stages, while further reductions in *t* make it efficient to locate stages according to their comparative costs.

3.2 The snake: Nash equilibrium:

Finally, we consider the equilibrium in which parts of the chain are controlled by independent firms, each of which is assumed to control a segment of length δ.[10] The technological relationship between stages of the snake mean that the location decision of any one stage will have consequences for other stages both upstream and downstream. For example, suppose in Figure 7 that is it cost reducing to move A + D + B collectively to S, but not to move any of these separately (given that others are in N, and if $\tau(\bar{z}_1)t$ and $\tau(\underline{z}_2)t$ are large). Coordinated decision taking looks at the whole range, but independent decision will not be able to average out profitable and unprofitable moves, so off-shoring is blocked by unprofitable of moving to S for firms in range D.

We briefly outline the implications of uncoordinated decision taking for the cases studied in the preceding sub-section. The question is; when is it profitable for one stage to relocate, given the location of other stages, and assuming that transactions between stages occur at unit cost? Suppose first that upstream is labour-intensive. The most upsteam firm controls interval [0, δ] of the value chain, and the cost difference from producing in S rather than N is

$$\Delta C(0,\delta) = \int_0^{\delta} [c(z)-1]dz + \tau(\delta)t \qquad (30)$$

i.e. the production cost difference plus off-shoring costs incurred, given that downstream stages are in N. If $c(0) < 1$ and the most downstream activity incurs no off-shoring costs $\tau(\delta)t = 0$, then even a very small agent will off-shore. But if $\tau(\delta)t > 0$ then off-shoring will be achieved only if the agent is large enough (δ big enough) to capture a range of production cost savings sufficient to cover the 'fixed cost' of off-shoring.

---

[10] At N's factor prices any interval of length δ accounts for fraction δ of the total value added in the product.

If this first move is profitable, do other firms follow? If the first *z* of the snake is located in S, then relocation from N to S by a firm occupying [*z*, *z* + δ] gives cost difference

$$\Delta C(z, z+\delta) = \int_z^{z+\delta} [c(z)-1]dz + \tau(z+\delta)t - \tau(z)t \tag{31}$$

Off-shoring costs are incurred on output (stage *z* + δ) but are saved on inputs (stage z). If δ→0 then linear approximation to the right hand side of (31) means that relocation is cost reducing so long as $c(z)-1+\tau'(z)t < 0$. This is just the gradient of total costs, $\partial U(z)/\partial z$ (equation (24)), so independent behaviour will cause relocation to occur so long as this derivative is negative. The Nash equilibrium and cost minimisation therefore coincide providing for the case where the first order condition characterises cost minimisation, and there is no discontinuity, $\lim_{z\to 0} \tau(z) > 0$ so that the first stage can move moved.

What about deviations from the downstream end? The most downstream firm controls interval [1- δ, 1], and the cost difference from producing in S rather than N is

$$\Delta C(0, \delta) = \int_{1-\delta}^{1} [c(z)-1]dz + \tau(1-\delta)t + \tau(1)t \tag{32}$$

i.e. the production cost difference plus transport costs incurred, given that both the market and upstream stages are in N. If this move by the most downstream firm is profitable, is it profitable for others to move from N to S? This is the decision of a firm in interval *z*, *z* – δ, where upstream firms are in N and downstream firms are in S. The move would save production costs, save costs on off-shoring output downstream, and incur costs on importing inputs. The cost difference is therefore

$$\Delta C(z, z+\delta) = \int_{z-\delta}^{z} [c(z)-1]dz - \tau(z)t + \tau(z-\delta)t \tag{33}$$

If δ → 0 then (33) is just the slope of $D(\tilde{z})$ in the open interval (0, *z*). But (32) includes the jump in costs associated with incurring off-shoring costs on both inputs and outputs. It will be satisfied only if δ is large enough to yield cost savings on a relatively wide range of production stages. Failing this, the equilibrium will not achieve global cost minimisation. Relocation is 'blocked', because the most downstream firm faces the cost penalty of locating away from both the market and upstream suppliers. The blockage can be overcome only if a coordinated move of a sufficiently large number of stages can be achieved.

What this suggests is that, given some distribution of parameters across sectors of the economy, off-shoring is more difficult to achieve in sectors where downstream activities are labour intensive and/or light, i.e. are the apparent candidates to benefit from off-shoring. Coordination failure may be acute for a downstream production stage sandwiched between a market in N and suppliers in N, whereas upstream stages can peel off more easily.

4. Concluding comments

It is a commonplace to say that globalization is associated with the fragmentation, unbundling and off-shoring of production. But how does this occur given the complexity of actual production processes or 'value chains'? No general results can be derived because of this complexity. Tasks with quite different characteristics may be 'adjacent' to each other in the value chain. For example, a footloose task (one with low communication and shipping costs) might be either capital- or labour-intensive, or tasks with similar factor intensity may have very different off-shoring costs. Nevertheless, this paper makes a stab at establishing results that provide some insights.

The first distinction is between snakes and spiders, in which 'adjacency' is quite different. Snakes are production processes where a physical entity follows a linear process with value added at each stage. Spiders are many limbed, with parts from different sources coming together in one place for assembly. In practise the two are combined: spiders might be attached to any part of a snake, an d multiple snakes might join into a spider.

How does activity in these two different models relocate from N to S (given a market in N and low labour costs in S) as trade and coordination costs fall? Production costs induce labour intensive (lower cost in S) parts to relocate, but this is moderated by the benefits of co-location with other stages of the production process. Discontinuous change and 'overshooting' can arise because of the role of node elements of production. In the spider, assembly is a nodal point, and when assembly relocates so do a wide range of parts. Some of these move against their comparative production costs in order to get the benefits of co-location, and then move back if off-shoring costs fall further. It is possible that assembly as a whole moves against its comparative production costs in order to get the benefits of co-location with labour intensive parts. If it is relatively capital intensive assembly may locate in N at very high and very low off-shoring costs, and in S at intermediate values of these costs. In the snake, the most downstream product is a node, as all upstream stages have to pass through this and thence to market. In the case where downstream is labour intensive this creates a barrier to off-shoring which, it if occurs at all, will be discontinuous, involving movement of a wide range of downstream activities together once a critical threshold is reached. Once again, co-location effects may induce overshooting, so stages move from N to S and then back again as off-shoring costs fall.

The comments above apply to the case when location decisions are made by a single agent seeking to minimise total costs. Equilibrium outcomes – where firms take decisions given the location of other firms – might not minimise total costs, as co-location forces generate coordination failures. It is of course possible to think of ways to overcome coordination failure, such as acquisition of various stages by a single firm, or leadership in a multi-stage game. However, in the simplest case coordination failure creates multiple equilibria and has the effect of blocking relocation. Firms are reluctant to relocate if they are not sure that they will be followed, and this is a source of inefficiency. Off-shoring is likely to occur more slowly than is socially optimal.

The tension between factor costs and co-location is, we think, central to a micro-analysis of off-shoring. Combining spiders and snakes or working with less continuous technologies (e.g. one stage of the snake has a spike in its factor intensity) creates many more situations where collocation benefits may cause parts to move against their apparent comparative advantage, or where equilibrium location is blocked by particularly hard to move stages. Interactions and collocation benefits may also extend beyond ‘adjacent’ stages in the value chain if, for example, there are synergies in design and product specifications. There are also the wider benefits of agglomeration, such as shared labour skills and knowledge spillovers. The research challenge is to produce further regularities from the mass of possible cases in this rich and important topic.

Appendix.

With $a_S > a_N$ the double intersection of then $A(t)$ and $R(t)$ can only occur if average parts' production costs in S are less than in N, $1 > \left(\bar{b} + \underline{b}\right)/2$, so $R(t)$ has non-positive gradient. Since $R(t) = 1$ at $t = \bar{b} - 1$, a sufficient condition is that $A(t) \leq 1$ at $t = \bar{b} - 1$. This condition is satisfied if $\frac{\bar{b} + \underline{b}}{2} < 1 - \frac{1}{2}\left\{1 - \frac{a_S - a_N}{\bar{b} - 1}\right\}$.

Figures are constructed with the following parameter values:

Figures 3 and 5: $a_S = 0.2$, $a_N = 0.3$, $\alpha = 0.5, \underline{b} = 0.4, \bar{b} = 1.4$, implying $t^A = 0.332$.

Figures 4 and 6: $a_S = 0.25$, $a_N = 0.2$, $\alpha = 0.275, \underline{b} = 0.4, \bar{b} = 1.2$, implying $t^A = 0.224, 0.536$

Figure 8, 9: $c(z) = g_0 + g_1 z$: $\tau(z) = \theta_0 + \theta_1 z + \theta_2 z^2$

Figure 8: $g_0 = 0.45$, $g_1 = 1$: Figure 9: $g_0 = 1.45$, $g_1 = -1$.

Case 1: $\theta_0 = 0$, $\theta_1 = 0.4$, $\theta_2 = 0$.

Case 1: $\theta_0 = 0.1$, $\theta_1 = 0.1$, $\theta_2 = 0$.

Case 1: $\theta_0 = 0.2$, $\theta_1 = -0.1$, $\theta_2 = 0$.

Case 1: $\theta_0 = 0$, $\theta_1 = 1.075$, $\theta_2 = -1$.

REFERENCES

Ando, M. and Kimura, F. (2005). The formation of international production and distribution networks in East Asia. In T. Ito and A. Rose (Eds.), International trade (NBER-East Asia seminar on economics, volume 14), Chicago: The University of Chicago Press. First version, NBER Working Paper 10167.

Arndt, S.W., Kierzkowski, H., (Eds.) (2001). Fragmentation: New Production Patterns in the World Economy. Oxford University Press, Oxford.

Baldwin, R. and F. Robert-Nicoud, (2010) "Trade-in-goods and trade-in-tasks: An Integrating Framework," NBER Working Papers 15882.

Dixit, A.K. and G. M. Grossman (1982) 'Trade and Protection With Multistage Production' *The Review of Economic Studies*, Vol. 49, No. 4 , pp. 583-594

Fujita, M. and J. Thisse (2006) 'Globalization and the evolution of the supply chain: who gains and who loses?' International Economic Review, 47, pp 811-836

Grossman, G. M., Rossi-Hansberg, E., (2008). Trading Tasks: A Simple Theory of Off-shoring, American Economic Review, 98, pp 1978-97.

Hanson, G. H., Mataloni, R. J., Slaughter, M., (2001). Expansion strategies of U.S. multinational firms, in: Rodrik, D., Collins, S. (Eds.), Brookings Trade Forum 2001, 245-282.

Hanson, G. H., R. J. Mataloni, and M. J. Slaughter, (2005). Vertical production networks in multinational firms, Review of Economics and Statistics 87, 664-678.

Harrigan, J. and A.J. Venables (2006) 'Timeliness and agglomeration', Journal of Urban Economics, 59, 300-316.

Helpman, E., (1984). A Simple Theory of Trade with Multinational Corporations. Journal of Political Economy 92, 451-471.

Helpman, E., (2006). Trade, FDI and the organization of firms', Journal of Economic Literature, XLIV, 589-630.

Helpman, E., Krugman, P.R., (1985). Market Structure and Foreign Trade, MIT Press, Cambridge MA.

Hummels, D. (2001). "Time as a Trade Barrier." Department of Economics, Purdue University, Lafayette, IN.

Hummels, D., Ishii, J. and Yi , K-M., (2001), The Nature and Growth of Vertical Specialization in World Trade. Journal of International Economics 54, 75-96.

Hummels, D., D. Rapoport, and K.-M. Yi, (1998). Vertical Specialization and the Changing Nature of World Trade. . Federal Reserve Bank of New York Economic Policy Review 4, 79-99.

Hummels, D., J. Ishii and Kei-Mu Yi, (1999). "The nature and growth of vertical specialization in world trade," Staff Reports 72, Federal Reserve Bank of New York.

Johnson, R. C. and Noguera, G. (2009) "Accounting for Intermediates: Production Sharing and Trade in Value Added". Unpublished, http://sites.google.com/site/robjohnson41/.

Jones, R. W., Kierzkowski, H., (2001). A Framework for Fragmentation, in: Arndt, S.W. Kierzkowski, H., (Eds), Fragmentation: New Production Patterns in the World Economy, Oxford University Press, Oxford.

Jones, R.W., (2000). Globalization and the Theory of Input Trade, MIT Press, Cambridge.

Kimura, F., Y. Takahashi and K. Hayakawa, (2007). "Fragmentation and parts and components trade: Comparison between East Asia and Europe," The North American Journal of Economics and Finance, Elsevier, vol. 18(1), pages 23-40, February.

Leamer, E.,and J. Levinsohn (1995). International trade theory; the evidence, in: Grossman, G., Rogoff, K., Handbook of International Economics, Vol. 3. North-Holland, Amsterdam., pp1339–1394

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

Markusen,J.R. and A.J. Venables (2007) 'Interacting factor endowments and trade costs: a multi-country, multi-good approach to trade theory', Journal of International Economics, 73, 333-354

Ng, F. and A. Yeats, (1999). Production sharing in East Asia; who does what for whom and why. World Bank Policy Research Working Paper 2197, Washington DC.

Venables, A.J. (1996), 'Equilibrium locations of vertically linked industries', International Economic Review, 37, 341-59.

Yeats, A., (1998). Just how big is global production sharing World Bank Policy Research Working Paper 1871, Washington DC

Yi, K-M., (2003). Can Vertical Specialization Explain the Growth of World Trade? Journal of Political Economy 111, 52-102.