



**DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES**

**RELOCATING THE VALUE CHAIN: OFF-SHORING AND
AGGLOMERATION IN THE GLOBAL ECONOMY**

Richard Baldwin and Anthony J. Venables

Number 544
March 2011

Manor Road Building, Oxford OX1 3UQ

1/March/2011

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
Graduate Institute
Geneva
Switzerland

A.J. Venables
Department of Economics
Manor Road
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), Jonkoping (European Regional Science Association) and Denver (ASSA).

1. Introduction:

An ever finer division of labor 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) and the current phase of globalisation as the second unbundling (production stages unbundled across nations). The first unbundling was triggered by steam power and subsequent advances in transportation technology. Transportation of goods and people is important in the second unbundling, but the trigger was advances in information and communications technology that massively lowered the cost of organising complex activities over distances. 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, as there are still costs to off-shoring stages of production. The cost and unpredictable delays involved in intercontinental shipping and travel of technicians and managers still matter, particularly with just-in-time management techniques. 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). 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 belts). In auto-production unbundling is sometimes ‘modularised’, with parts coming together for assembly in a ‘module’ which is then shipped whole to the final assembly plant (Frigant and Lung 2002). For example, US car seats are typically assembled in a plant within one hour of the auto assembly plant; there are virtually no US imports of complete car seats, but imports of parts of car seats are around \$5bn pa, of which half come from China (Klier and Rubinstein 2009).

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 2001, 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).

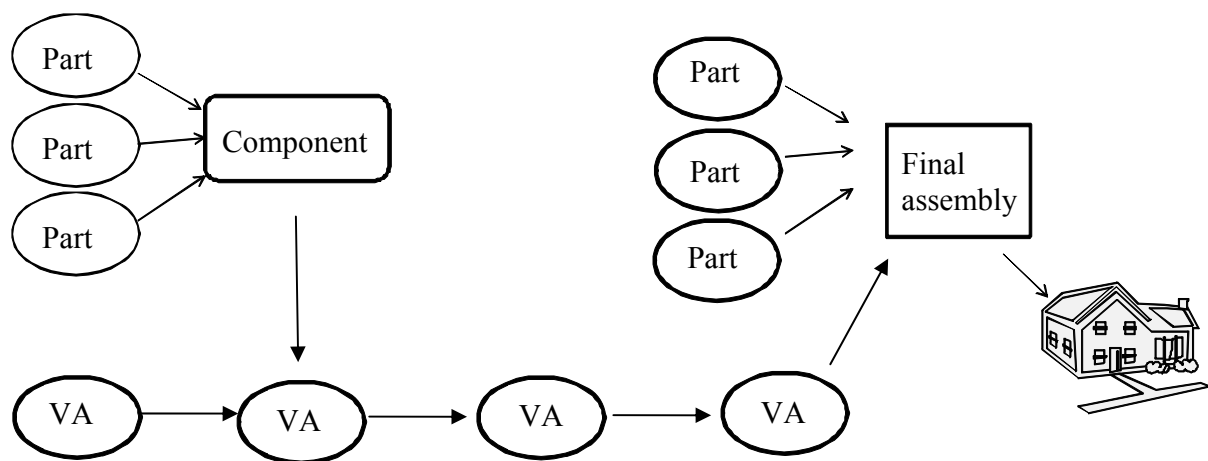
Analytical work has taken a variety of approaches. Much of the focus has been on taking simple characterisations of the technology of unbundling and drawing out the general equilibrium implications for trade and particularly for wages (Yi 2003, Grossman and Rossi-Hansburg 2008, Markusen and Venables 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.¹ 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. These may be movements within a country (or within a plant), or may be unbundled movements between plants in different countries. There are two quite

¹ Some of the management literature discusses off-shoring in terms of the ‘tasks structured by the underlying project architecture’ (Sako 2005).

different configurations. One we refer to as the spider: multiple limbs (parts) coming together to form a body (assembly), which may be the final product itself or a component (such as a module in the auto-industry). The other is the snake: the good moving in a linear manner from upstream to downstream with value added at each stage.² Most production processes are complex mixtures of the two. Cotton to yarn to fabric to shirts is a snake like process, but adding the 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.³ These off-shoring costs create centripetal forces binding related stages together. Firms seek to be close to other firms with which they transact, and the form of this depends on the engineering of the product; it is different for snakes (each stage linked to an upstream and a

² Dixit and Grossman (1982) undertake a general equilibrium analysis of the snake. Levine (2010) and Costinot et al (2010) study the implications of a snake in which mistakes can occur at each stage.

³ And costs associated with length and variability of time in transit, Harrigan and Venables (2006).

downstream stage) than for spiders (each part linked just to assembly, but assembly linked to many parts). 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 equilibrium outcomes when stages are controlled by independent decision takers. In the cost-minimising case the extent of off-shoring and the volume of trade are non-monotonic functions of off-shoring costs, because of interaction between comparative costs and co-location forces. In the equilibrium, co-location and agglomeration forces mean that outcomes are not necessarily cost minimising since coordination failures obstruct moves towards efficiency. There will be multiple equilibria and locational hysteresis. These issues have arisen in earlier work on inter-industry linkages (e.g. Venables 1996, Fujita et al 1999) but in a stylised Dixit-Stiglitz structure of symmetric firms producing differentiated products. The present paper moves significantly beyond this work, with a more detailed focus on the heterogeneity between different parts of the chain, and the interplay between these stages.

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 questions we address are: which stages are produced in N, which are produced in S, and how does this balance change during a process of globalisation?

⁴ We can relax this to an assumption that more of the final demand is in N than in S, but the assumption that all is in N un-clutters the analysis.

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 \bar{b} and we assume $\underline{b} < 1 < \bar{b}$, 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 ‘labor-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 $\psi(y)$ units of each part $y \in Y$. 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.

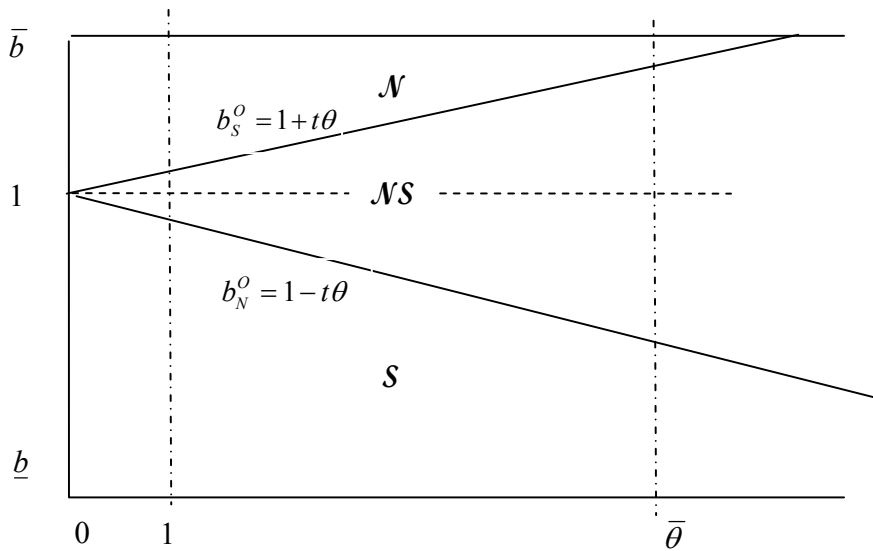
When a part’s production is spatially separated from assembly a per-unit off-shoring cost of $t\theta(y)$ is incurred, this representing shipping costs and a wider set of communication, coordination and trade costs. The cost is the product of a parameter t capturing the overall level of off-shoring costs, 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 $[1, \bar{\theta}] \times [\underline{b}, \bar{b}]$ and each part can be represented as a point

⁵ We call this as assembly of the final product although it could be assembly of a component sold in N.

$\{b(y), \theta(y)\}$ 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 b , the cost of production in S. In the bottom region, \mathcal{S} , the converse is true: these are parts which are relatively labor-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 $1 > b + t\theta$. The two lines dividing these sets can be thought of as ‘off-shoring thresholds’. Values of b on the upper line will be labelled $b_S^O = 1 + t\theta$, since parts on this line can be produced and delivered to assembly in S at the same costs, regardless of whether they are produced in N or 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



\mathcal{N} : set of parts produced in N when assembly in N or in S.

\mathcal{S} : set of parts produced in S when assembly in N or in S.

\mathcal{NS} : set of parts produced in N if assembly in N, in S if assembly is in S.

With this as set up, we now determine the location of assembly and associated location of parts. We first discuss this in 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 of 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 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.

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

The left hand side of the inequality is assembly costs in N (equal to a_N) plus the costs of parts; types in sets \mathcal{N} and \mathcal{NS} are produced in N at unit cost 1, whilst those in set \mathcal{S} are produced in S at cost $b(y)$ and also incur off-shoring costs $t\theta(y)$ to reach the assembler. The right hand side has the assembler operating in S at primary factor cost a_S with parts in sets \mathcal{NS} and \mathcal{S} produced in S and those in set \mathcal{N} produced in N. Off-shoring costs are incurred on parts in set \mathcal{N} and also on the assembled product that has to be shipped to the market in N.⁶

Inequality (1) can be rearranged as

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

⁶ If there is final demand in both N and S, then αt is the difference between the average cost of reaching markets when assembly is in S as compared to when it is in N.

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 \mathcal{MS} 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 \mathcal{M} , while saving off-shoring costs on parts from set \mathcal{S} .

By inspection of Figure 2 and equation (2) we see that, as $t \rightarrow 0$, all activities locate where their primary factor costs are lowest. The set \mathcal{MS} disappears, so parts locate according to the value of $b(y)$ relative to unity, while assembly locates in S if $a_S < a_N$, and in N if $a_S > a_N$. At the other extreme, as $t \rightarrow \infty$, assembly and production of all parts takes place in N, where final demand is. The sets \mathcal{M} and \mathcal{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 $t \rightarrow 0$ and off-shoring costs are decisive as $t \rightarrow \infty$, 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 \mathcal{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 \mathcal{MS} , just leaving parts in \mathcal{M} to be produced in N. However, further reductions in t enlarge set \mathcal{M} (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, or there may alternatively be an intermediate range of t in which assembly takes place in S, relocating parts in set \mathcal{MS} as it does so. This happens if a high proportion of parts are labor intensive so that set \mathcal{M} is small. It is then efficient to undertake assembly and production of parts in sets \mathcal{M} and \mathcal{MS} in 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 off-shoring 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.⁷ 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

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

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 parts in sets \mathcal{N} and \mathcal{NS} , while saving those in set \mathcal{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 \mathcal{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

⁷ This can be rationalised by a contestability assumption ensuring that parts' producers make zero profits. We relax the assumption in section 2.3.

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

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 since 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.

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_N^o , b_S^o , so set \mathcal{N} is parts with b values greater than b_S^o , set \mathcal{S} parts with b less than b_N^o , and set \mathcal{NS} is parts with $b \in [b_N^o, b_S^o]$.

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

$$C(b_N^o : N) = a_N + \int_{b_N^o}^{\bar{b}} \psi(b) db + \int_{\underline{b}}^{b_N^o} (b+t) \psi(b) db \quad (5)$$

$$C(b_S^o : S) = a_S + \alpha t + \int_{b_S^o}^{\bar{b}} (1+t) \psi(b) db + \int_{\underline{b}}^{b_S^o} b \psi(b) db,$$

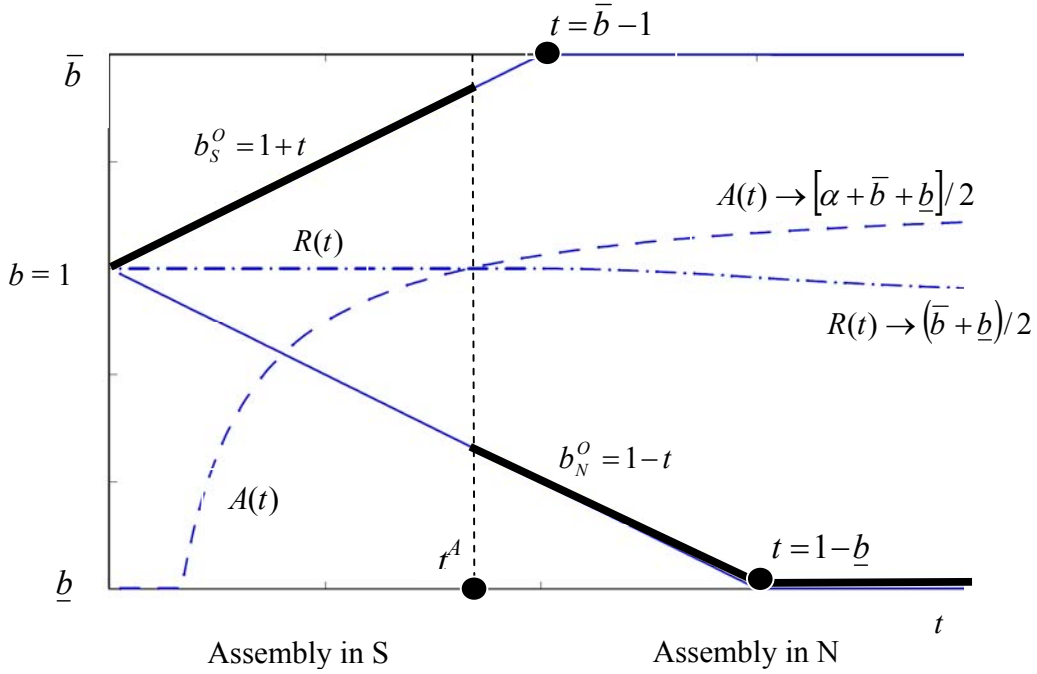
these equations corresponding to the two sides of inequality (1). The threshold values b_N^o , b_S^o minimise the costs of supply in each case, and are therefore given by first order conditions

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

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 parameter 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 $[\underline{b}, b_N^o]$ form the set \mathcal{S} ; those in $[b_S^o, \bar{b}]$ form set \mathcal{NS} ; and those in $[b_S^o, \bar{b}]$ form set \mathcal{N} .

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



Where will the assembler locate, 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

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

This expression depends on t , and to analyse further it is convenient to split it 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 (with $a_S < a_N$) and Figure 4 ($a_S > a_N$); it is cost minimising to locate assembly in S (equation (8) is positive) if $R(t)$ lies above $A(t)$.

If t is less than $\min[1 - \underline{b}, \bar{b} - 1]$ then $R(t) = 1$ since the off-shoring thresholds are simply $b_S^o = 1 + t < \bar{b}$ and $b_N^o = 1 - t > \underline{b}$ (and hence $b_S^o + b_N^o = 2$). If t exceeds $\bar{b} - 1, 1 - \underline{b}$ then 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 $t \rightarrow \infty$. The dependence of $A(t)$ on t is clear from equation (10). As $t \rightarrow 0$, so $A(t)$ tends to minus infinity if $a_S < a_N$, or to plus infinity if $a_S > a_N$. As $t \rightarrow \infty$, 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 characterize potential intersections of the curves, as tabulated below:

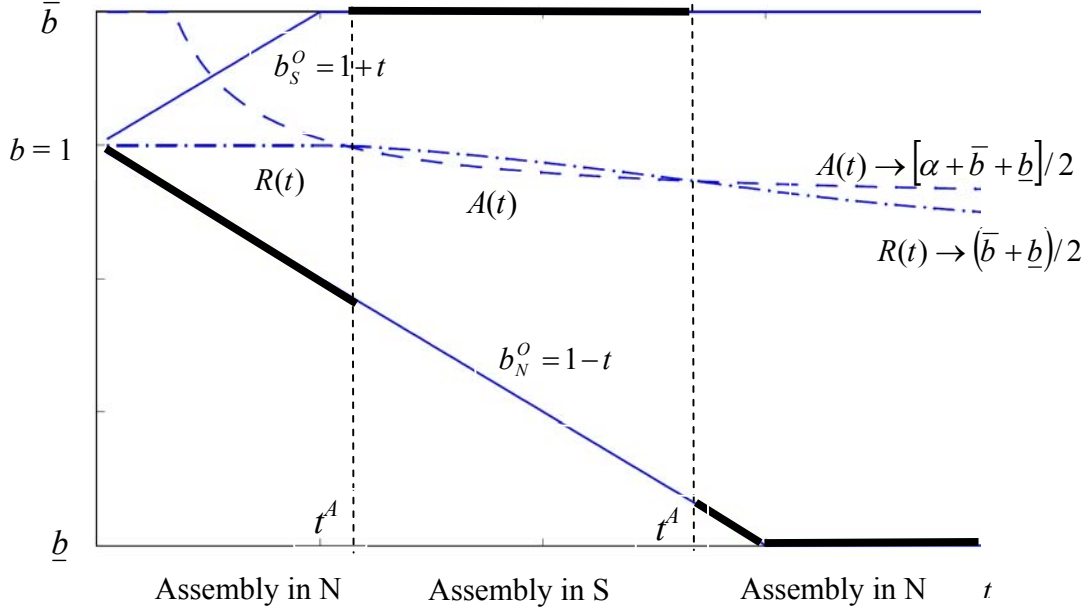
Table 1: Location of assembly

	$t \rightarrow 0$	$t \rightarrow \infty$	Intersections
$a_S < a_N$	$A(t) \rightarrow -\infty$ $< R(t) = 1$ Assembly in S	$A(t) \rightarrow [\alpha + \bar{b} + \underline{b}]/2$ $> R(t) \rightarrow (\bar{b} + \underline{b})/2$ Assembly in N	1 intersection, t^A
$a_S > a_N$	$A(t) \rightarrow +\infty$ $> R(t) = 1$ Assembly in N	$A(t) \rightarrow [\alpha + \bar{b} + \underline{b}]/2$ $> R(t) \rightarrow (\bar{b} + \underline{b})/2$ 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 cost minimising 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 these lines. Initially all production is in N, then declining t is associated with a slow migration of low b parts (labor 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, $[\underline{b}, \bar{b}]$ 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 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 locate assembly in S in order to best use low cost parts producers in S; 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 $(\underline{b} + \bar{b})/2 < 1$ 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 section 2.2) assumed that parts differ in their production costs, b , but all face the same off-shoring costs, $t\theta(y) = t$. What if there is a correlation between the two, for example labor-intensive goods having high off-shoring costs? This can be captured by making off-shoring costs a function of b , such as $t[1+h(b)]$, with $h(1) = 0$ and $[1+h(b)] > 0$ for all $b \in [\underline{b}, \bar{b}]$.⁸ If $h'(b) < 0$ then parts that are cheaper to produce in S have high off-shoring costs, i.e. $h(b) > 0$ for $b < 1$, and $h(b) < 0$ for $b > 1$. We look at 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

⁸ In terms of the formulation of section 2.1 and Figure 2, the set Y is a line, but it is no longer vertical.

minimum, are

$$\begin{aligned}\partial C(b_N^O : N) / \partial b_N^O &= \psi(b_N^O) [b_N^O - (1 - t[1 + h(b_N^O)])] = 0 \\ \partial C(b_S^O : S) / \partial b_S^O &= \psi(b_S^O) [b_S^O - (1 + t[1 + h(b_S^O)])] = 0.\end{aligned}\tag{12}$$

At $t = 0$ the threshold values of b are unity, as before. With $h'(b) < 0$ the thresholds are rotated downwards (clockwise) from this point since $[1 + h(b)] < 1$ for $b > 1$ and $[1 + h(b)] > 1$ for $b < 1$. This rotation is intuitive since it means that, conditional on the location of assembly, fewer parts are produced in S.

What about the location of the assembler? Inspection of (11) indicates that having $h'(b) < 0$ 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 the cost difference $C(b_N^O : N) - C(b_S^O : S) = 2t[R(t) - A(t)]$, the function $R(t)$ is unchanged (equation 9) and $A(t)$ (equation 10) becomes

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

(see appendix). 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 labor intensive parts have the effect of *increasing* off-shoring to S.

In summary, if parts that are cheaper to produce in S have relatively high off-shoring costs then, given the location of assembly, there is less production in S. But the range of t values for which assembly takes place in S is increased. The intuition is that it is more expensive to access S's cost advantage 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 $[b_N^O, b_S^O]$.

Economies of scope in off-shoring:

A further possibility is that off-shoring costs depend on the number or range of parts that are produced in a particular location. Economies of scope may mean that as more parts are off-shored so total off-shoring costs increase less than proportionately. Such economies 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 $t.g(\cdot)$, where $g(\cdot)$ 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, this additional effect reduces costs by $tg(b_N^o - \underline{b})$, and symmetrically if assembly is in S. Total costs are then,

$$C(b_N^o : N) = a_N + \int_{b_N^o}^{\bar{b}} \psi(b)db + \int_{\underline{b}}^{b_N^o} (b+t)\psi(b)db - tg(b_N^o - \underline{b}) \quad (14)$$

$$C(b_S^o : S) = a_S + \alpha t + \int_{b_S^o}^{\bar{b}} (1+t)\psi(b)db + \int_{\underline{b}}^{b_S^o} b\psi(b)db - tg(\bar{b} - b_S^o).$$

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,

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

If the midpoint of the support $[\underline{b}, \bar{b}]$ is unity then Figures 3 and 4 would be symmetric around $b = 1$, so $g(\bar{b} - b_S^o) = g(b_N^o - \underline{b})$ and economies of scope have no impact on the location of the

assembler. However, if the midpoint of $[\underline{b}, \bar{b}]$ is less than unity (as Figures 3 and 4 are actually drawn) then, at each t , $g(\bar{b} - b_S^o) < g(b_N^o - \underline{b})$. $A(t)$ is 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 ($\theta(y) = 1$) and $\psi = 1$. The equilibrium location of parts producers is given by the sets \mathcal{N} , \mathcal{NS} and \mathcal{S} , defined by off-shoring thresholds b_S^o, b_N^o , as before. Inequality (3), the condition for assembly to be in N, becomes

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

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

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

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; on this line 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, $a_S < a_N$. 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, 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 ($a_S < a_N$)

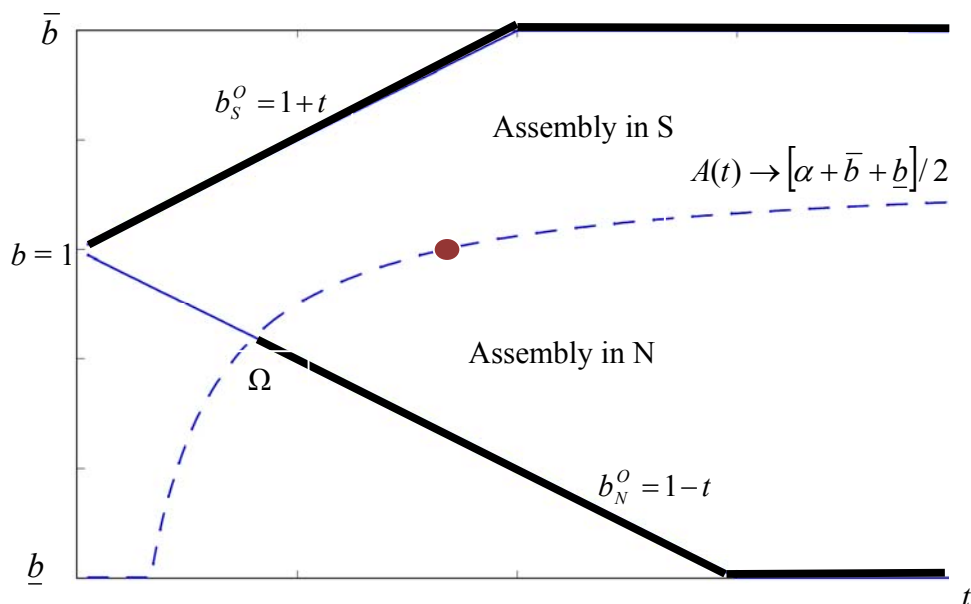
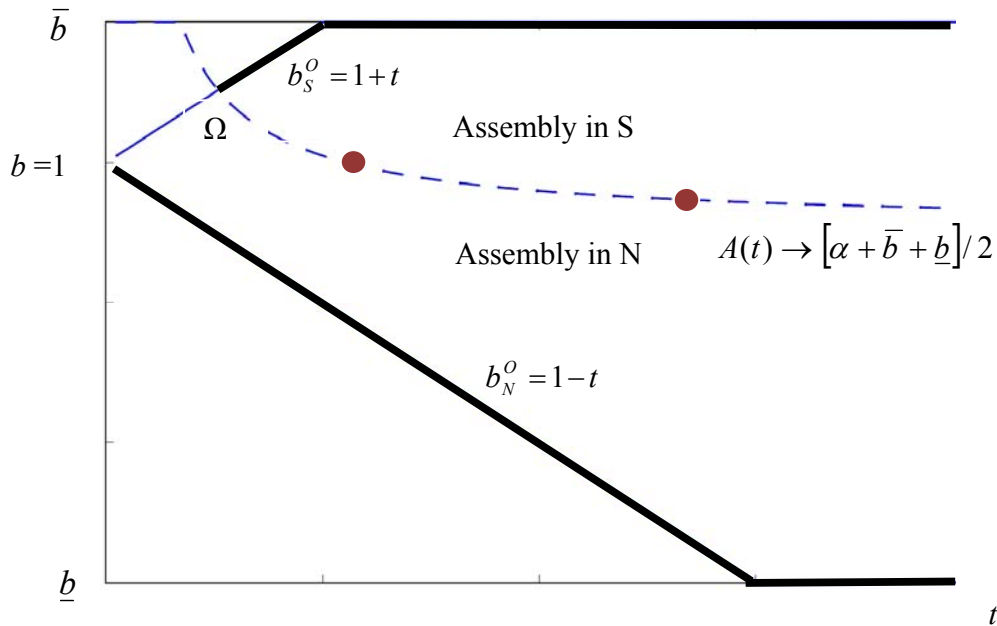


Figure 6 gives the case where primary factor costs of assembly are higher in N than in S. As illustrated assembly in N is an equilibrium at all values of t ; starting from this position, offshoring of parts takes place in a continuous manner but assembly never moves. However, there is, for t greater 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 6 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 equilibrium has assumed that parts are sold to the assembler at unit cost, suggesting that this can be supported by contestability – the potential entry of parts producers disciplines firms to price at cost. 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 decisions 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 we suppose that access to technology and entry is easy) but assume 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 the producer in S and potential entrants in N. 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 by p_N , (respectively p_S), given by:

$$p_N = b + \gamma(1 - (b + t)), \quad p_S = b + \gamma(1 + t - b). \quad (18)$$

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, what is 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 that the Nash bargain is efficient). Assembler's costs are

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

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. Alternatively, 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

$$C(b_S^O : S) = a_S + \alpha t + (1+t)(\bar{b} - b_S^O) + \int_{\underline{b}}^{b_S^O} (b + \gamma(1+t-b))db. \quad (20)$$

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

$$\begin{aligned} C(b_N^O : N) - C(b_N^O : S) &= a_N - a_S - \alpha t - t(\bar{b} - b_N^O) + \int_{\underline{b}}^{b_N^O} t(1-2\gamma)db \\ &= a_N - a_S - t \left[\alpha + (\bar{b} - b_N^O) - (b_N^O - \underline{b})(1-2\gamma) \right] \end{aligned} \quad (21)$$

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

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

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 to S moves the assembler into an environment with less competition in the supply of parts; higher γ therefore further postpones off-shoring, amplifying the inefficiency we saw in the previous section.

2.4. 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 range of parts. These movements may 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 labor 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 fail 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 $z \in (0,1)$ 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 are set equal to unity in N and denoted $c(z)$ in S, so a low c value denotes a stage with high labor 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, and the unit cost is $\tau(z)t$, with $\tau(z) > 0$ for $z \in (0,1]$. As before, t is a parameter common to all stages representing the overall technology of off-shoring. $\tau(z)$ captures the way in which off-shoring costs vary with z ; this may be increasing as value is added, or decreasing as weight is lost, e.g. in refining or

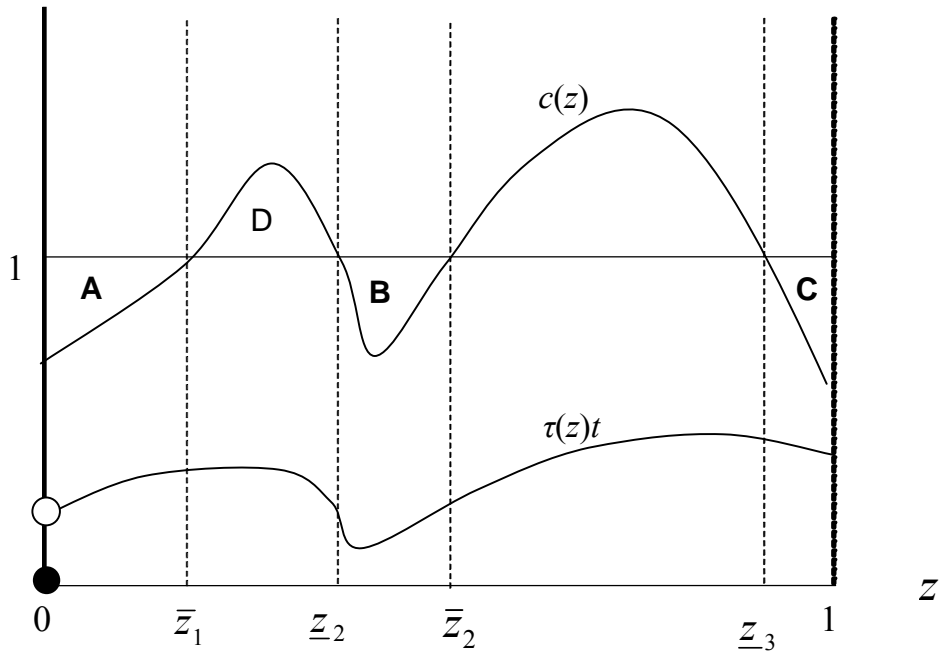
smelting a natural resource. We will look at both cases.

Figure 7 illustrates production and off-shoring costs along the snake, and we start with informal discussion of this figure 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.⁹ As illustrated, stages in the ranges marked A, B, C are ‘labor 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-sourcing costs are incurred if no stages are produced in S), while it is possible that $\lim_{z \rightarrow 0} \tau(z) > 0$ (moving even the smallest stage to S incurs off-shoring costs). Interval B has low production costs if undertaken in S but, if 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 inputs are imported and output exported. 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 ($\tau'(\bar{z}_2) < 0$). 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 production 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.

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.

⁹ Drawn to be continuous, although not monotonic.

Figure 7: Costs along the snake



3.1 The snake: cost minimisation:

What is the efficient division 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 for all z , with $c(0) < 1$ and $c(1) > 1$, so that upstream is labor intensive and downstream capital intensive. The cost minimisation problem can be set up as choice of critical value \hat{z} , upstream of which ($z \leq \hat{z}$) production takes place in S, while downstream stages $z > \hat{z}$ take place in N. \hat{z} is chosen to minimise total costs, denoted $U(\hat{z})$, and given by

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

If $\hat{z} = 0$ then all production takes place in N and total costs are unity. If $\hat{z} > 0$ the second equation applies, 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.¹⁰ We assume that the functions $c(z)$ and $\tau(z)$ are twice differentiable, so first and second derivatives with respect to \hat{z} are

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

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

A number of issues arise in finding 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 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 are off-shored. The second rules out situations with sharply falling marginal off-shoring costs, like 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,

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

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, for a product that loses weight as it is processed, $\tau'(z) < 0$, a lower t reduces the range of activities in S, enabling the comparative advantage of N in downstream stages to operate over a wider range of the snake.

¹⁰ The objective has to be written in these two parts because of the discontinuity in $\tau(z)$ noted above.

Third, the local minimum may not be a global minimum because of the discontinuity at zero, $\tau(0) = 0$ but $\lim_{z \rightarrow 0} \tau(z) > 0$. In this case values of U at zero and at the local minimum must be evaluated and compared. The comparison indicates that, for very high t it is certainly cheaper to do all production 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 the second order condition fails if $\tau''(\hat{z}) < 0$ and t is large enough. The cost minimising outcome may then 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 $\tau(z) = \theta_0 + \theta_1 z + \theta_2 z^2$. Figure 8 gives four cases, differing in coefficients $\theta_0, \theta_1, \theta_2$ (see appendix for values). 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} on the vertical, with stages below the lines produced in S.

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 $\lim_{z \rightarrow 0} \tau(z) > 0$ 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 .

Figure 8a: Upstream off-shoring

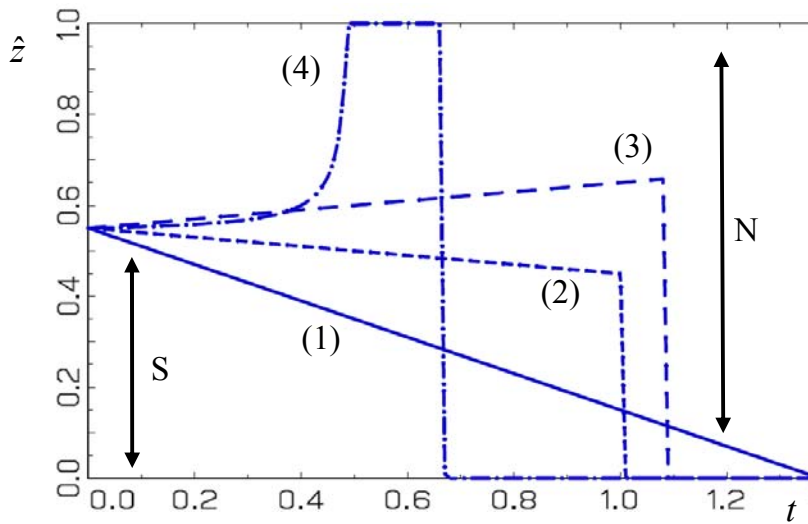
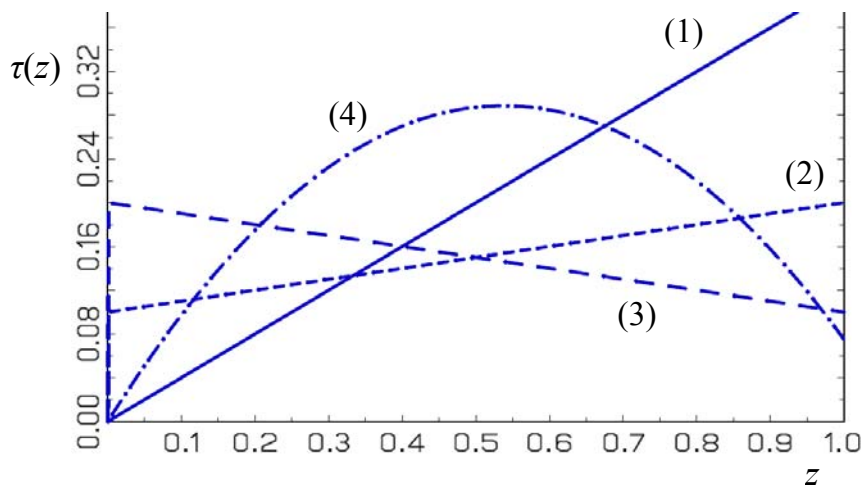


Figure 8b: Functions $\tau(z)$



Finally, case (4), in which off-shoring costs are (strongly) concave, with $\tau''(z) < 0$. At high t the second order condition fails, so an interior solution $\hat{z} \in (0,1)$ is not efficient. As illustrated, $\tau(1) > \tau(0)$ so at high t the snake is entirely in N, while at intermediate values it switches to S. At still lower values of t the second order condition comes to hold, so the snake splits with upstream activities off-shored to S and downstream taking place in N. This is then a quite dramatic example of overshooting. At high off-shoring costs all parts co-locate, initially in

N and then in S. At low enough off-shoring costs there is an interior division of the snake between countries.

This analysis assumed that production in S occurs (if at all) for upstream stages, validating this with S having a cost advantage in these stages, $c'(z) > 0$, $c(0) < 1$ and $c(1) > 1$. We now reverse comparative costs supposing that $c'(z) < 0$, $c(0) > 1$ and $c(1) < 1$, so S has cost advantage in downstream stages. 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:

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

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

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* shipping costs, $\tau(z)t$ plus $\tau(1)t$. As before, there is a discontinuity where off-shoring commences, with $D(\tilde{z})$ jumping upward as \tilde{z} goes below unity. Derivatives, for $\tilde{z} \in (0,1]$, are

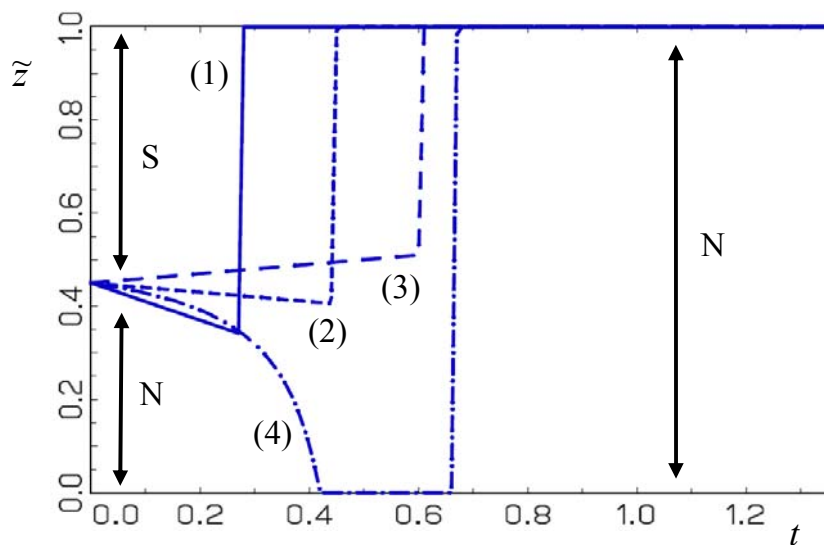
$$\frac{\partial D(\tilde{z})}{\partial \tilde{z}} = 1 - c(\tilde{z}) + \tau'(\tilde{z})t . \quad (28)$$

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

Outcomes are illustrated on Figure 9, with $\tau(z)$ taking the same four shapes as in Figure 8b, but $c(z)$ now linearly decreasing. Stages above each line (high z) are produced in S. Three points stand out. First, if $\tau(1)t > 0$, 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.

Figure 9: Downstream off-shoring



Outcomes are illustrated on Figure 9, with $\tau(z)$ taking the same four shapes as in Figure 8b, but $c(z)$ now linearly decreasing. Stages above each line (high z) are produced in S. Three points stand out. First, if $\tau(1)t > 0$, 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.

Pulling together the results of this section we see that, even with the simplest case of linear $c(z)$ and quadratic $\tau(z)$, reductions in the off-shoring cost parameter t gives a complex set of outcomes. Off-shoring is like a fixed cost, so there are discontinuities. Reductions in t can

either increase or decrease the range of stages produced in S. And there is ‘overshooting’ with stages moving and then moving back as t falls.

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 .¹¹ If δ is small it is apparent that off-shoring which reduces costs collectively, over an interval of the snake, will not necessarily take place. For example, in Figure 7 it may be cost reducing to move A + D + B collectively to S, but not to move any of these separately (given that other stages 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 the unprofitability 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 labor-intensive. The most upstream firm controls interval $[0, \delta]$ of the value chain, and the cost difference from producing in S rather than N is

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

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.

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 $[z, z + \delta]$ gives cost difference

¹¹ At N’s factor prices any interval of length δ accounts for fraction δ of the total value added in the product.

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

Off-shoring costs are incurred on output (stage $z + \delta$) but are saved on inputs (stage z). If $\delta \rightarrow 0$ then (31) implies 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 for the case in which the first order condition characterises cost minimisation and in which there is no discontinuity, $\lim_{z \rightarrow 0} \tau(z) > 0$ so that the first stage can move.

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

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

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 $z, z - \delta$, 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

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

If $\delta \rightarrow 0$ then (33) is just the slope of $D(\tilde{z})$ in the open interval $(0, z)$, similar to the upstream off-shoring case.

Comparing the upstream and downstream cases, the main point to note is the double incidence of off-shoring costs in (32) as compared to (30); downstream off-shoring has to both import its inputs and export its output, while upstream off-shoring only does the latter. Off-shoring is therefore possible only if δ is large enough to yield cost savings on a relatively wide range of production stages. Failing this, relocation is ‘blocked’, because the most downstream

firm faces the cost penalty of locating away from both the market and upstream suppliers. 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 labor intensive so are the apparent candidates to benefit from off-shoring. Coordination failure may be acute for a labor intensive downstream production stage sandwiched between a market in N and suppliers in N. The blockage can be overcome only if a coordinated move of a sufficiently large number of stages can be achieved.

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 off-shoring costs) might be either capital- or labor-intensive, or tasks with similar factor intensity may have very different off-shoring costs. Nevertheless, this paper makes establishes a number of 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, and 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 labor costs in S) as trade and coordination costs fall? Production costs induce labor 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 labor intensive parts. If assembly is relatively capital intensive it 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 labor intensive this creates a barrier to off-shoring which, if it 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 which have 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. 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.¹² We think that the approach of this paper suggests a research program to investigate these issues.

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 co-location 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 co-location 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 labor skills and knowledge spillovers. The research challenge is to produce further regularities from the mass of possible cases in this rich and important topic.

¹² Some greenfield auto-assembly plants (such as VW’s Resende plant and Ford’s Bahia plant) are accompanied by construction of a ‘supplier park’ in which capital costs are met by the final assembler (Sako 2005).

Appendix.

Table 1: 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, $1 > (\bar{b} + \underline{b})/2$, 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 $\frac{\bar{b} + \underline{b}}{2} < 1 - \frac{1}{2} \left\{ 1 - \frac{a_S - a_N}{\bar{b} - 1} \right\}$.

Equation (13): Equation (7) becomes

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

Integrating and splitting into $A(t)$ and $R(t)$ gives equation (13).

Figures: 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: $c(z) = g_0 + g_1z$: $\tau(z) = \theta_0 + \theta_1z + \theta_2z^2$

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 2: $\theta_0 = 0.1$, $\theta_1 = 0.1$, $\theta_2 = 0$.

Case 3: $\theta_0 = 0.2$, $\theta_1 = -0.1$, $\theta_2 = 0$.

Case 4: $\theta_0 = 0$, $\theta_1 = 1.075$, $\theta_2 = -1$.

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.
- 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.
- Costinot, A., J. Vogel and S. Wang (2010) 'Vertical specialisation and the interdependence of nations' draft, MIT
- 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.
- Frigant, V. and Y. Lung (2002) 'Geographical proximity and supplying relationships in modular production', *International Journal of Urban and Regional Research*, 26.4 pp742-55
- 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.
- 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.
- Klier, T.H. and J. Rubinstein (2009) 'Imports of Intermediate parts in the auto-industry; a case study' <http://www.upjohninst.org/measurement/klier-rubenstein-final.pdf>
- Levine, D. (2010) 'Production chains', NBER wp 16571
- 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.
- Sako, M. (2005) 'Governing automotive supplier parks: leveraging the benefits of outsourcing and co-location', processed, SBS, Oxford.
- Venables, A.J. (1996), 'Equilibrium locations of vertically linked industries', International Economic Review, 37, 341-59.
- Yi, K-M., (2003). Can Vertical Specialization Explain the Growth of World Trade? Journal of Political Economy 111, 52-102.