

NBER WORKING PAPER SERIES

OPTIMAL INTEGRATION STRATEGIES
FOR THE MULTINATIONAL FIRM

Gene M. Grossman
Elhanan Helpman
Adam Szeidl

Working Paper 10189
<http://www.nber.org/papers/w10189>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2003

We acknowledge with thanks the support of the National Science Foundation (SES 9904480 and SES 0211748) and the US-Israel Binational Science Foundation (2002132). The views expressed herein are those of the authors and not necessarily those of the National Bureau of Economic Research.

©2003 by Gene M. Grossman, Elhanan Helpman, and Adam Szeidl. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Optimal Integration Strategies for the Multinational Firm
Gene M. Grossman, Elhanan Helpman, and Adam Szeidl
NBER Working Paper No. 10189
December 2003
JEL No. F23, F12, L22

ABSTRACT

We examine integration strategies of multinational firms that face a rich array of choices of international organization. Each firm in an industry must provide headquarter services from its home country, produce intermediate inputs, and assemble the intermediate goods into final products. Both production of intermediate goods and assembly can be performed at home, in another “Northern” country, in the low-wage “South,” or in several of these locations. We study the equilibrium choices of firms that differ in productivity (and thus size), focusing on the role of industry characteristics such as the fixed costs of foreign subsidiaries, the cost of transporting intermediate and final goods, and the share of the consumer market that resides in the South in determining optimal integration strategies.

Gene M. Grossman
Department of Economics
300 Fisher Hall
Princeton University
Princeton, NJ 08544
and NBER
grossman@princeton.edu

Elhanan Helpman
Department of Economics
Littauer Center
Harvard University
Cambridge, MA 02138
and NBER
ehelpman@harvard.edu

Adam Szeidl
Department of Economics
Littauer Center
Harvard University
Cambridge, MA 02138
szeidl@fas.harvard.edu

1 Introduction

The globalization process of recent years has been expressed in the growth of many types of international transactions, but few more salient than the expansion in the activity of multinational firms. The growth rate of sales by foreign affiliates of multinational corporations outpaced the growth of exports of goods and non-factor services by almost seven percent per year from 1990 to 2001. Gross product by all foreign affiliates accounted for an estimated eleven percent of world GDP in 2001, while exports by these affiliates represented an estimated 35 percent of total world trade (UNCTAD, 2002).

Multinational firms have pursued a multitude of strategies for international expansion, as described in the *World Investment Report* (UNCTAD, 1998) and cited by Yeaple (2003). Firms have opened foreign affiliates to perform activities ranging from R&D to after-sales service, and including production of parts and components, assembly, and wholesale and retail distribution, among others. Some firms procure parts from subsidiaries in many countries and assemble them in a single location. Others concentrate production of parts in one place and assemble final products in several plants located close to their customers. Still others erect an integrated plant in a low-wage country and use it to serve consumers around the globe. The motives for foreign direct investment (FDI) are similarly diverse, but the potential for factor-cost savings, for transportation-cost and trading-cost savings, and for the realization of economies of scale seem to be among the primary inducements.

The theory of international trade and foreign direct investment traditionally has distinguished two forms of multinational activity based on alternative reasons why a firm might opt to locate production or other activities abroad (see, for example, Markusen [2002, pp.17-20]). *Vertical* multinationals are firms that geographically separate various stages of production. Such fragmentation of the production process typically is motivated by cost considerations arising from cross-country differences in technologies or factor prices. For example, Helpman (1984) and Helpman and Krugman (1985) model multinational firms that maintain their headquarters in one country but manufacture output in another in order to conserve on production costs. In contrast, *horizontal* multinationals are firms that replicate most or all of the production process in several locations. These multi-plant firms often are motivated by potential savings of transport and trading costs. In the models developed by Markusen (1984), Brainard (1997) and Markusen and Venables (1998, 2000), for example, firms with headquarters in a home country produce final output in plants that serve consumers in each of two national markets.

The distinction between vertical and horizontal FDI is clear enough when there are two countries and two production activities, namely headquarter operations and “manufacturing.” But with more countries and more stages of production, some organizational forms do not fit neatly into either of these categories. For example, a multinational firm might manufacture goods in a foreign subsidiary and sell the output primarily in third-country markets; Ekholm

et al. (2003) term such activity “export-platform FDI.” Or a firm might perform intermediate stages of production in one country to save on production costs and subsequent stages in several plants to conserve on transport costs. Yeaple (2003) follows the *World Investment Report* in referring to this as a “complex integration strategy.” Feinberg and Keane (2003) report that, in their sample of U.S. multinationals with affiliates in Canada, only 12 percent of the firms have negligible intra-firm flows of intermediate goods and thus can be considered to be purely horizontal multinationals, while only 19 percent of the firms have intra-firm flows of intermediate goods in only one direction, which would make them purely vertical multinationals. The remaining 69 percent of firms are what they call “hybrids”; i.e., firms that are pursuing more complex integration strategies. Similarly, Hanson et al. (2001) describe the rich patterns of FDI they find in their data pertaining to operations by U.S. multinationals and their foreign affiliates. They document and analyze the roles played by foreign affiliates as export platforms, as producers adding value to inputs acquired from their U.S. parents, and as wholesale distributors in foreign markets. Based on their analysis of data for the 1990’s, Hanson et al. conclude that “the literature’s benchmark distinction between horizontal and vertical FDI does not capture the range of strategies that multinationals use.”

Both Yeaple (2003) and Ekholm et al. (2003) examine theoretically the determinants of firms’ choices among a limited set of integration strategies that includes an option for FDI that is neither purely horizontal nor purely vertical. Yeaple studies a model with two identical “Northern” countries and a third, “Southern” country in which firms headquartered in one of the Northern countries need two produced inputs to assemble differentiated final goods. One component can be produced more cheaply in the North, the other in the South. Shipping entails an “iceberg” transport cost that is a similar proportion of output for intermediate goods as for final goods. All consumption of the differentiated final goods takes place in the North. In this context, Yeaple compares the profitability of four integration strategies: (i) a “national firm” that produces both of the components in the same Northern country as where its headquarters are located; (ii) a “vertical multinational” that produces one component in the South and the other in the firm’s home country; (iii) a “horizontal multinational” that maintains integrated production facilities (that produce both components) in both Northern countries, and (iv) a “complex multinational” that produces one component in the South and the other in both Northern countries. In Yeaple’s model of symmetric producers, all firms adopt the same integration strategy in equilibrium. Yeaple shows how the viability of the four different organizational forms depends on factor-price differentials, shipping costs, and the fixed costs of establishing subsidiaries in the North and South.

Ekholm et al. (2003) also study a setting with two similar Northern countries and a single Southern country. Theirs is a duopoly model, with one firm headquartered in each country in the North. Each of these firms must produce an intermediate good in its home country but may assemble their final output in one or more plants located in any or all of the countries.

Thus, each firm chooses among four options: (i) a national firm that conducts all activities at home, (ii) a purely horizontal multinational that assembles in both Northern countries; (iii) a pure export platform, with all assembly in the South; and (iv) a hybrid multinational, with assembly in both the home country and the South. Like Yeaple, Ekholm et al. examine how the organizational choices reflect transport costs, the relative cost advantage of the South, and the fixed costs associated with foreign investment.

Our concerns in this paper are somewhat similar to those of Yeaple (2003) and Ekholm et al. (2003), but we aim to shed light on the determinants of integration strategy when firms face a richer array of choices. Our goal is to provide a reasonably general analysis in which a variety of different complex integration strategies can emerge in equilibrium. In our model, as with the others, there are three countries; namely, two, symmetric Northern countries that we call “East” and “West” and a low-wage country that we call “South.” In contrast to the earlier papers, we allow for consumption of the differentiated products produced by integrated firms in all three locations. Thus, the relative size of the Southern market becomes an important parameter in our analysis. The firms that produce differentiated products must perform two production activities besides their headquarter services; they first must produce intermediate goods and then must assemble the intermediates into a final product. Either production of intermediate goods, or assembly, or both may be separated geographically from a firm’s headquarters, and a firm may perform these activities in one or several locations. As in Yeaple (2003), there are interesting complementarities that link the location decisions.

We assume in our analysis that the prospective labor costs of producing intermediate goods and assembling them are lower in the South than in the North. A firm must bear a fixed cost for each plant it operates abroad to produce intermediate goods and a (possibly different) fixed cost for each foreign subsidiary that assembles final goods. Both intermediate goods and final goods may be costly to trade, and the cost of transporting the two types of goods (relative to the value of output) need not be the same. The key parameters that we use to describe an industry are the sizes of the transport costs for intermediate and final goods, the relative size of the fixed costs for different types of subsidiaries, and the share of the consumer market that resides in the South.

We also allow for heterogeneity among the firms in an industry. Following Melitz (2002) and Helpman et al. (2003), we assume that each entrant into an industry draws a productivity level from a known distribution. By the time that firms make their decisions about integration strategy, they have learned about their own potential productivity levels. In equilibrium, firms with different productivity levels may make different choices about their organizational form. Thus, our model can account for the coexistence of a variety of forms in the same industry, in keeping with the evidence reported by Hanson et al. (2001) and Feinberg and Keane (2003). Moreover, our analysis draws a link between the size of a firm and its equilibrium integration strategy. In principal, these predictions can be subjected to empirical

scrutiny.

The remainder of this paper is organized as follows. In Section 2, we develop our model of firms that must choose where to produce intermediate goods and where to assemble final products. The firms in an industry share similar fixed costs of opening foreign subsidiaries, similar costs of shipping components, and similar costs of shipping final goods. They face symmetric demands but differ in their potential productivity. In Section 3, we analyze the equilibrium integration strategies that emerge in the absence of transport costs. In this simple case we are able to develop intuition about the sorting of firms by productivity level and show how the parameters describing fixed costs and the relative size of the South affect the choices of organizational form. In Section 4, we introduce transportation costs for final goods and consider the full range of possible costs from low to high. Again we examine how different parameters describing industry conditions color the equilibrium choices by firms with different productivity levels. Section 5 contains a discussion of some interesting cases that arise when intermediate goods too are costly to transport. Section 6 concludes.

2 The Model

We seek a simple setting in which firms face a choice between performing activities at home and engaging in foreign direct investment (FDI) to conserve on either production costs or trading costs. We also need to distinguish between “assembly activities”—those that result in a finished product ready for sale to consumers—and “intermediate activities”—those that can be performed in any location so long as the output later is transported to the place of assembly. For this, we develop a model with three countries and two stages of production. Following Ekholm et al. (2003) and Yeaple (2003), we assume that one of the countries (‘South’) has low production costs and a relatively small market for the goods produced by the integrated firms, while the other two (‘East’ and ‘West’, together comprising the ‘North’) have larger markets, higher wages, and are fully symmetric.

Households consume goods produced by $J + 1$ industries. One industry supplies a homogeneous good under competitive conditions. The others manufacture differentiated products. Consumers share similar preferences that can be represented by the utility function

$$U = x_0 + \sum_{j=1}^J \frac{1}{\mu_j \alpha_j} X_j^{\mu_j}, \quad 0 < \mu_j < 1, \quad (1)$$

where x_0 is consumption of the homogeneous good and X_j is an index of consumption of the differentiated outputs of industry $j \in \{1, \dots, J\}$. The consumption index for industry j is a CES aggregate of the amounts consumed of the different varieties. That is,

$$X_j = \left[\int_0^{n_j} x_j(i)^{\alpha_j} di \right]^{1/\alpha_j}, \quad 0 < \alpha_j < 1, \quad (2)$$

where $x_j(i)$ is consumption of the i^{th} variety of industry j and n_j is the measure (number) of varieties in that industry. With this utility function, the elasticity of substitution between any pair of goods produced by industry j is $1/(1 - \alpha_j)$. We assume that $\alpha_j > \mu_j$, so that the brands in a given industry substitute more closely for one another than they do for the outputs of a different industry.

We distinguish the countries in several ways. First, firms in the North are more productive than those in the South in producing the homogeneous good. This creates a gap between Northern and Southern equilibrium wages. We assume that one unit of labor is needed to produce one unit of the homogeneous good in East or West, but that $1/w > 1$ units of labor are needed to produce one unit of the good in South. We also assume that the homogeneous good is produced in the equilibrium in all three countries and take this good to be the numeraire. Then $w^E = w^W = 1 > w^S = w$, where w^ℓ is the wage in country ℓ . Second, the sizes of the markets for differentiated products may differ; we denote by M^ℓ the number of households in country ℓ who consume differentiated products, and assume that $M^E = M^W = M^N$.¹ Finally, we assume that firms can enter as producers of differentiated products only in the Northern countries and that these firms must locate their headquarters in their country of origin.

Entry into industry j requires h_j units of local labor in East or West. With this fee, entrants acquire the design for a differentiated product and learn their productivity level, θ . Productivity levels in industry j are independent draws from a cumulative distribution function, $G_j(\theta)$. A firm in industry j with productivity θ produces final output according to the production function $\theta F_j(m, a)$, where m is the quantity of a specialized, intermediate input and a is the level of assembly activity. The intermediate goods can be produced apart from the assembly activity, but if so, the intermediates must be shipped to the place of assembly before a final good can be produced. The location of assembly determines the (pre-shipment) location of the final good.

We take $F_j(\cdot)$ to be an increasing and concave function with constant returns to scale and an elasticity of substitution between m and a no greater than one. Let $c_j(p_m, p_a)$ denote the unit cost function dual to $F_j(m, a)$, where p_i is the effective price of input i in the place of assembly (including delivery costs). Then $c_j(p_m, p_a)/\theta$ is the per-unit variable cost of production in this location for a firm with productivity θ .

A firm in industry j that separates the production of intermediate inputs from the location of its headquarters bears an extra (fixed) cost of g_j units of home labor for communication and governance. These costs are the same for a firm that produces the intermediates in the other Northern country as for one that produces them in the South. Similarly, a firm that engages in FDI in assembly incurs an extra fixed cost of f_j units of home labor no matter

¹We do not necessarily associate the number of consumers of differentiated products with the size of a country's population. There may be some consumer who lack sufficient income to consume these products and who instead concentrate their purchases on the homogeneous good.

where the assembly takes place. Iceberg transportation costs may exist for both intermediate inputs and final goods. Specifically, a firm in industry j must ship $\tau_j \geq 1$ units of the intermediate good to deliver one unit of the good to a distant place of assembly and $t_j \geq 1$ units of the final good to deliver one unit of the good to a distant place of consumption.

We assume that the manufacture of one unit of an intermediate good requires one unit of local labor in the place of production and that one unit of assembly activity requires one unit of local labor in the place of assembly. With these assumptions, the South enjoys a comparative advantage both in assembly and in production of intermediate goods relative to production of the homogeneous good x_0 .²

It is now straightforward to calculate the variable cost to a firm in industry j of delivering one unit of the final good to a given market by means of alternative integration strategies. Consider for example a firm in East with productivity θ that wishes to deliver final goods to consumers in West. Such a firm would pay $t_j c_j(1, 1)/\theta$ per unit to produce and assemble the good at home (including the cost of shipping to West), whereas it would pay $t_j c_j(w, w)/\theta$ per unit to conduct all production and assembly activity in South. Still another possibility would be to produce intermediates in South and perform assembly in West, thereby avoiding the transport cost for final goods. The variable cost associated with this strategy would be $c_j(\tau_j w, 1)/\theta$ per unit, considering the cost of shipping the intermediates from South to West.

3 Zero Transport Costs

We begin with the case in which intermediate and final goods can be shipped between countries at zero cost. It is helpful to examine this simple case first, because it highlights the trade-off between the fixed costs of FDI and the variable-cost savings that can be achieved by performing certain activities in the low-wage South (as in Helpman et al. [2003]), as well as the complementarities between FDI decisions for different stages of development (as in Yeaple [2003]).

In what follows, we consider firms in a particular industry j and omit the subscript j from the variables and parameters of interest. We focus on the variation across firms in productivity levels, as indexed by θ . The firms under consideration may have their headquarters in East or West. Since these two countries are fully symmetric, it is more convenient to refer to H , the home country of the firm in question, and R , the “other” Northern country in which the firm will sell its output. This means, of course, that if $H = E$, $R = W$; and if $H = W$, $R = E$.

With zero transport costs, the integrated firm never opts to produce its intermediate goods

²We have also examined situations with different production structures that admit a comparative advantage for the South in one of the activities undertaken by the integrated firms. For small comparative advantage in one of these activities, our results are unaffected. Larger degrees of comparative advantage modify our result in fairly intuitive ways.

Table 1: Fixed and Per-Unit Variable Costs

production m	assembly a	fixed cost	per-unit variable cost
in H	in H	0	$c(1, 1) / \theta$
in H	in S	f	$c(1, w) / \theta$
in S	in H	g	$c(w, 1) / \theta$
in S	in S	$f + g$	$c(w, w) / \theta$

or conduct assembly operations in country R , because the variable costs are the same in R as in H and FDI imposes extra fixed costs. Moreover, a firm has no reason to undertake a given activity in two locations, as this would impose extra governance costs without conserving on any transport costs. Thus, four integration strategies remain for consideration: production of intermediates might take place either in H or S and assembly might occur either in H or S . Table 1 shows the fixed and per-unit variable costs associated with each of the four strategies. The fixed costs indicated are those *extra* costs that result from operating one or more foreign subsidiaries.

The first row depicts a strategy of home production. With this strategy, the firm serves the foreign markets in R and S with exports from its home assembly plant. As is clear, this strategy provides for a low fixed cost, as it avoids the extra costs associated with FDI. But the per-unit variable cost is high, because factor prices are higher in E or W than in S . The following two rows depict strategies of “partial globalization”; either intermediates are produced at home and assembled in South (second row), or vice versa (third row). These strategies yield intermediate levels of fixed and variable costs; they cannot be ranked vis-à-vis one another without further information about the cost function $c(\cdot)$ and the sizes of the fixed costs for the two types of foreign subsidiaries. With assembly in S , the firm exports intermediates from its home plant, and then exports finished goods from S to consumers in H and R . This means that the strategy combines elements of “vertical FDI” and what Ekholm et al. (2003) have termed “export-platform FDI.” With intermediates produced in S , there again is intra-firm trade, as well as exports of final goods from H to markets in R and S . The bottom row depicts a strategy of complete globalization, whereby all production activities are performed in the low-wage South. Here, fixed costs are highest, variable costs are lowest, and the markets in H and R are served by exports from South. With this strategy, there is no trade in intermediate goods.

We can readily compare the operating profits that a firm with productivity θ can achieve under the alternative strategies. Considering the form of consumer preferences in (1) and (2), every firm in the industry faces a demand function in market ℓ given by

$$x^\ell = \alpha^{-\alpha/(1-\alpha)} M^\ell \left(X^\ell \right)^{(\mu-\alpha)/(1-\alpha)} \left(p^\ell \right)^{-1/(1-\alpha)}, \quad (3)$$

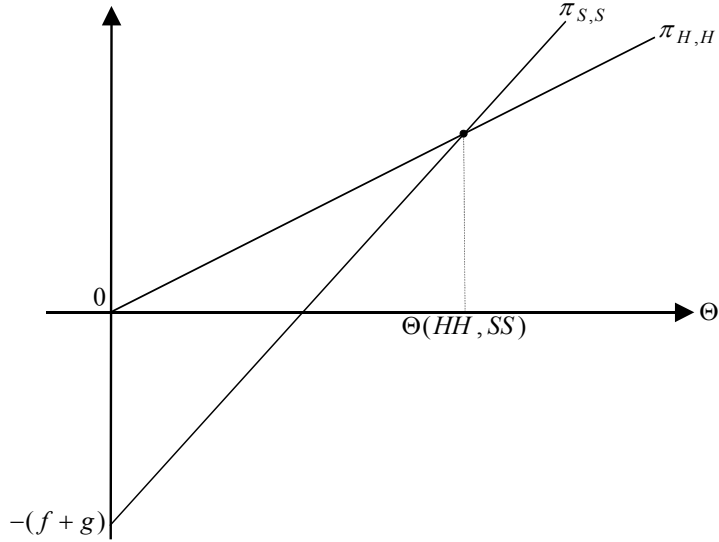


Figure 1: Profitability of Home Production and Complete Globalization

where X^ℓ is the aggregate consumption index for varieties in the industry in country ℓ and p^ℓ is the price it charges there. Each producer treats the aggregate consumption indexes as given. Therefore, it maximizes profits by charging a price in each market that is a multiple $1/\alpha$ of its per-unit variable cost of serving that market. Since the per-unit cost of serving each market is the same when transport costs are zero, so too are the optimal prices associated with a given strategy. It follows from the demand function in (3) that, for any strategy with an extra fixed cost of k and a per-unit variable cost of c/θ , the maximum attainable operating profits are

$$\pi = (1 - \alpha)\bar{Y}\Theta c^{-\alpha/(1-\alpha)} - k,$$

where $\Theta \equiv \theta^{\alpha/(1-\alpha)}$ is another measure of the firm's productivity and $\bar{Y} \equiv \sum M^\ell (X^\ell)^{(\mu-\alpha)/(1-\alpha)}$ is a measure of the size of the world market.

In Figure 1, we depict the operating profits attainable from home production (the top row in Table 1) and complete globalization in South (the bottom row in Table 1), for different levels of productivity Θ . These profits, which we denote by $\pi_{H,H}$ and $\pi_{S,S}$, are given by

$$\pi_{H,H} = \frac{(1 - \alpha)\bar{Y}\Theta}{C(1, 1)} \quad (4)$$

and

$$\pi_{S,S} = \frac{(1 - \alpha)\bar{Y}\Theta}{C(w, w)} - (f + g) \quad (5)$$

respectively, where $C(p_m, p_a) \equiv [c(p_m, p_a)]^{\alpha/(1-\alpha)}$ is a transformed measure of unit cost. The figure shows that firms with low productivity prefer home production whereas firms with

high productivity prefer FDI, in keeping with the findings of Helpman et al. (2003). The reason, of course, is that FDI offers the prospect of lower per-unit costs and higher fixed costs, and the potential to save on variable cost is most valuable to highly productive firms that anticipate producing high volumes of output.

Next consider the firm's option to locate only assembly operations in South, while producing intermediate goods in the home country. The potential operating profits from this integration strategy for a firm with productivity Θ are

$$\pi_{H,S} = \frac{(1-\alpha)\bar{Y}\Theta}{C(1,w)} - f. \quad (6)$$

If we were to add $\pi_{H,S}$ to Figure 1, it would have an intercept between those of $\pi_{H,H}$ and $\pi_{S,S}$ and a slope steeper than $\pi_{H,H}$ but less steep than $\pi_{S,S}$. Thus, if locating only assembly in South is to be viable at any productivity level, this strategy must be at least as profitable as concentrating both activities in either location at the productivity level labelled $\Theta(HH,SS)$ in the figure. But this requires³

$$\frac{g}{f} \geq \frac{C(1,1)}{C(w,w)} \left[\frac{C(1,w) - C(w,w)}{C(1,1) - C(1,w)} \right]. \quad (7)$$

Leaving this strategy aside for a moment, the firm also has the option to produce intermediate goods in South and assemble final goods at home. This strategy offers a firm with productivity Θ operating profits of

$$\pi_{S,H} = \frac{(1-\alpha)\bar{Y}\Theta}{C(w,1)} - g. \quad (8)$$

Again, the intercept and slope are intermediate between those for the two lines shown in Figure 1, and viability of the strategy requires that it be at least as profitable as the other two at $\Theta = \Theta(HH,SS)$. This in turn requires

$$\frac{g}{f} \leq \frac{C(w,w)}{C(1,1)} \left[\frac{C(1,1) - C(w,1)}{C(w,1) - C(w,w)} \right]. \quad (9)$$

From (7) and (9) we conclude that no firm will separate the production of its intermediate goods from its assembly operations when

$$\frac{C(w,w)}{C(1,1)} \left[\frac{C(1,1) - C(w,1)}{C(w,1) - C(w,w)} \right] < \frac{g}{f} < \frac{C(1,1)}{C(w,w)} \left[\frac{C(1,w) - C(w,w)}{C(1,1) - C(1,w)} \right].$$

Our assumption that the elasticity of substitution between intermediates and assembly in the production of final goods is no greater than one ensures that the upper limit in this

³To derive this condition, we calculate $\Theta(HH,SS)$ as the value of Θ that equates $\pi_{H,H}$ and $\pi_{S,S}$, and then compare $\pi_{H,S}$ and $\pi_{H,H}$ at $\Theta = \Theta(HH,SS)$.

string of inequalities exceeds the lower limit.⁴ It follows that there always exist a range of values of g/f for which neither assembly in South and production of intermediates at home, nor production of intermediates in South with assembly at home is optimal for any firm, regardless of its productivity level.

Suppose now that the fixed costs of operating a foreign assembly operation are small relative to the fixed costs of operating a foreign plant to manufacture intermediate goods; i.e., g/f is large enough so that (7) is satisfied. Then a firm with productivity level at or near $\Theta(HH, SS)$ prefers to locate its assembly in South and manufacture intermediates at home to any other integration strategy. Figure 2 shows the operating profits $\pi_{H,S}$ (as well as $\pi_{H,H}$ and $\pi_{S,S}$) for this case. Clearly, firms with low productivity below $\Theta(HH, HS)$ conduct all operations at home, firms with intermediate productivity between $\Theta(HH, HS)$ and $\Theta(HS, SS)$ conduct only their assembly operations in South, and firms with high productivity above $\Theta(HS, SS)$ perform all of their production activities in South.

The case when the fixed cost of FDI in assembly is large relative to the fixed cost of FDI in intermediates is qualitatively similar. With g/f small enough so that (9) is satisfied, the line representing $\pi_{S,H}$ will cut $\pi_{H,H}$ at some relatively low productivity level $\Theta(HH, SH)$ that is to the left of $\Theta(HH, SS)$ in Figure 1, and will cut $\pi_{S,S}$ at some relatively high productivity level $\Theta(SH, SS)$ to the right of $\Theta(HH, SS)$ in the figure. Then firms with productivity between $\Theta(HH, SH)$ and $\Theta(SH, SS)$ will choose to produce their intermediates in the low-wage South while conducting assembly at home.

Our analysis can be used to highlight one form of complementarity that exists between a firm's decision to invest abroad at different stages of production. Compare, for example, a firm's decision whether to conduct assembly in South when $g = 0$ and $g = \infty$. In the first case, the fixed cost of FDI in intermediate goods is nil and so all firms produce their intermediates in South. In the second case, the cost of FDI for the production of intermediates is prohibitive and all firms produce their intermediates at home. When $g = 0$, the productivity level at

⁴It can be shown that

$$\frac{C(1, 1)}{C(w, w)} \left[\frac{C(1, w) - C(w, w)}{C(1, 1) - C(1, w)} \right] > \frac{C(w, w)}{C(1, 1)} \left[\frac{C(1, 1) - C(w, 1)}{C(w, 1) - C(1, 1)} \right]$$

if and only if

$$\frac{1}{C(w, w)} + \frac{1}{C(1, 1)} > \frac{1}{C(w, 1)} + \frac{1}{C(1, w)}$$

i.e., if and only if the function $1/C(\cdot)$ is supermodular. But $1/C(p_m, p_a) \equiv [c(p_m, p_a)]^{\alpha/(1-\alpha)}$ is supermodular if it is twice differentiable and

$$\frac{c(p_m, p_a) [\partial^2 c(p_m, p_a) / \partial p_m \partial p_a]}{[\partial c(p_m, p_a) / \partial p_m][\partial c(p_m, p_a) / \partial p_a]} < \frac{1}{1 - \alpha}.$$

The left-hand side of this inequality is the elasticity of substitution between m and a in the production of final goods, which is no greater than one by assumption. Therefore, the inequality holds for all positive values of α .

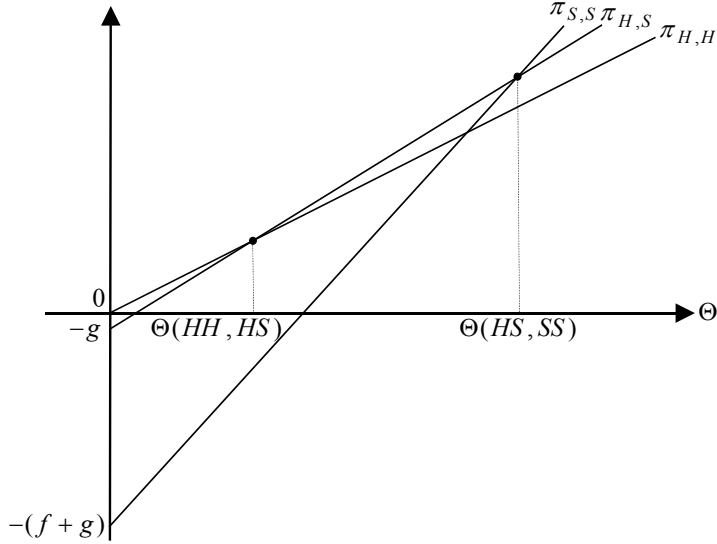


Figure 2: Partial globalization optimal for intermediate productivity levels

which a firm is indifferent between assembly at home and assembly in South is

$$\Theta(SH, SS) = \frac{f}{(1 - \alpha)\bar{Y} \left[\frac{1}{C(w,w)} - \frac{1}{C(w,1)} \right]},$$

whereas when $g = \infty$, the productivity level at which a firm is indifferent between assembly at home and assembly in South is

$$\Theta(HH, HS) = \frac{f}{(1 - \alpha)\bar{Y} \left[\frac{1}{C(1,w)} - \frac{1}{C(1,1)} \right]}.$$

In footnote 5 we showed that $1/C(\cdot)$ is supermodular when, as here, the elasticity of substitution between intermediates and assembly does not exceed one. It follows that $\Theta(SH, SS) < \Theta(HH, HS)$; i.e., firms will shift their assembly abroad for a wider range of productivity levels when they also produce their intermediates there than when they produce their intermediates at home.

Figure 3 can be used to summarize the arguments up to now. It shows the integration strategies chosen by different firms (as indexed by their productivity Θ) for different values of $g/(f + g)$. In drawing this figure, we hold constant the measure of world demand (as represented by \bar{Y}) and the combined fixed costs of FDI (as represented by $f + g$).⁵

For all strictly positive values of $g/(f + g)$, low-productivity firms in an industry perform all stages of production at home and export their final product to R and S . These firms

⁵If \bar{Y} increases or $f + g$ decreases, the boundaries between the various regions shift to the right, because these changes make it more profitable for a firm with given productivity to engage in FDI.

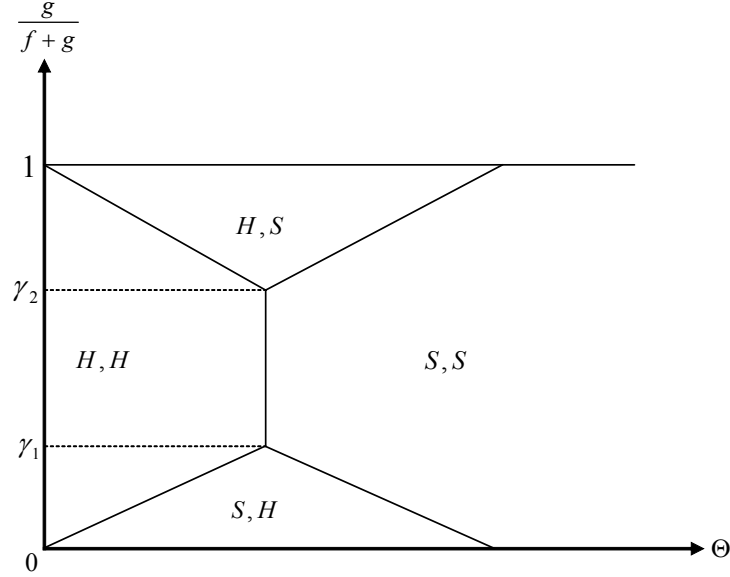


Figure 3: Integration Strategies for Different Productivities and Relative Fixed Costs

intend to produce relatively little output, so the savings in variable cost offered by FDI does not justify the higher fixed costs. Firms with intermediate levels of productivity may separate their production of intermediates from their assembly operations, depending on the relative sizes of f and g . If so, such firms will engage in intra-firm trade in addition to exporting final output either from their home assembly plant or from an export platform in South. Finally, high-productivity firms will perform all operations in the low-wage South so as to take greatest advantage of the low per-unit costs there.

Before leaving this section, we show how the model can be closed to construct an industry equilibrium. We define the envelope of the profit functions as

$$\pi(\Theta) = \max_{z_1 \in \{H,S\}, z_2 \in \{H,S\}} \pi_{z_1, z_2}(\Theta) ,$$

where $\pi(\Theta)$ is the operating profit earned by a firm with productivity Θ when it pursues its optimal integration strategy. Given the distribution of productivity levels $G(\theta)$, the free-entry condition can be written as

$$\int_0^\infty \pi \left[\theta^{\alpha/(1-\alpha)} \right] dG(\theta) = h .$$

Since the profit function is increasing in the measure of world demand \bar{Y} , which in turn is increasing in the aggregate consumption index X , the free-entry condition uniquely determines the industry value for X . All other industry variables, including the number of varieties and the cut-off points for each integration strategy can now be computed using this value of X .

4 Transport Costs for Final Goods

In this section, we allow for costly transport of final goods, while maintaining the assumption that intermediates can be shipped costlessly. For example, the intermediates may represent services that can be performed remotely and then moved electronically. This assumption implies that intermediates goods are only produced in one location.

When the transport of final goods is costly, relative market size may affect the location of assembly operations. Moreover, firms may engage in “horizontal FDI,” by, for example, assembling goods in more than one location. We focus here on the interaction between market size and FDI costs in determining a firm’s optimal integration strategy.

4.1 Low Transport Costs

The viable integration strategies vary with the size of transport costs. We begin with a case in which transport costs for final goods are reasonably small; in particular, we assume that

$$1 < t < \frac{c(1, 1)}{c(1, w)}. \quad (10)$$

When inequality (10) is satisfied, the variable cost of serving any market is minimized by assembly in South, no matter where the intermediate goods are produced. To see this, observe first that if the intermediates are produced in H or R , the cost of serving any market from an assembly plant in the North is at least $c(1, 1)$. But this exceeds the cost of serving the same market from the South, which is at most $tc(1, w)$. Next observe that if intermediates are produced in South, the per-unit variable cost of serving any market from an assembly plant in the North is at least $c(w, 1)$, while the per-unit cost of serving the same market from a plant in South is at most $tc(w, w)$. However, $c(w, 1)/c(w, w) > c(1, 1)/c(1, w)$,⁶ so inequality (10) ensures that $c(w, 1) > tc(w, w)$ as well.

Under the circumstances, a firm with headquarters in H will not conduct any activity in R . Intermediate goods are no less costly to produce in R than in H and can be shipped costlessly from one to the other. By producing these goods in R , the firm would needlessly incur an extra fixed cost of FDI. And if assembly is to be conducted outside of H , the delivered cost of serving any market from S are lower than the cost of serving the market from R , while the fixed cost of an assembly plant is the same in the two locations.

We can also rule out any integration strategy in which a given activity is performed in more than one location. If it is worthwhile for the firm to bear the fixed cost of opening a facility to manufacture intermediate goods in South, the firm produces all of its intermediates there to take full advantage of the low production costs. The same is true for assembly, considering

⁶Note that $c(1, 1)/c(1, w) < c(1, w)/c(w, w)$ if and only if $\log c(1, 1) + \log c(w, w) < \log c(1, w) + \log c(w, 1)$; i.e., if and only if $\log c(p_m, p_a)$ is submodular. But $\log c(p_m, p_a)$ indeed is submodular when the elasticity of substitution between m and a is less than one, because $\partial^2 \log c(p_m, p_a) / \partial p_m \partial p_a < 1$.

the reasonably low cost of shipping goods. It follows that each firm chooses one of four integration strategies; these are the same set of strategies that we considered in Section 3.

A firm's decision calculus is similar to that described in Section 3, except that now it must take into account the relative size of the market in South when deciding whether to open facilities there. We define $Y^\ell \equiv M^\ell (X^\ell)^{(\mu-\alpha)/(1-\alpha)}$ as a measure of market size in country ℓ and $\sigma \equiv Y^S/\bar{Y}$ as the share of the South in world demand for industry output.

We begin, as before, by comparing the profitability of performing all production activities at home with the profitability of performing all stages in South. Again, firms with low productivity prefer home production while those with high productivity prefer FDI in South. The productivity level $\Theta(HH, SS)$ at which a firm earns equal operating profits under the alternative integration strategies depends, as before, on the size of the fixed costs, $f + g$, and on the relative wage, w . Now it depends too on the shipping cost and on the relative size of South. If the typical market in the North is larger than the market in South (i.e., if $Y^E = Y^W = Y^N > Y^S$), then a larger shipping cost t diminishes the relative profitability of production in South, because the shipments from S to H and S to R , under this strategy are larger respectively than those from H to S and from H to R under the alternative strategy of home production. It follows that the larger is t , the greater is the productivity level that makes a firm indifferent between production in home and complete globalization in South.⁷ On the other hand, the larger is σ , the smaller is $\Theta(HH, SS)$, because the relative profitability of producing in South increases with the size of the market there due to the transport costs.

Again, as was the case with zero transport costs, a strategy of producing intermediate goods at home with assembly in South can be optimal for some firms only if the fixed cost of FDI in intermediate-good production is high relative to the fixed cost of a foreign assembly operation. And a strategy of producing intermediate goods in South with assembly at home can be optimal for some firms only if the fixed cost of FDI in intermediate-good production is relatively low compared to that for a foreign assembly operation. For intermediate values of g/f , all low-productivity firms produce their intermediates and conduct assembly at home while all high-productivity firms produce their intermediates and conduct assembly in South.

Figure 4 shows the optimal integration strategies for firms with different productivity levels Θ when g/f is large enough to satisfy

$$\frac{g}{f} \geq \frac{C(1, 1)}{C(w, w)} \left[\frac{C(1, w) - C(w, w)}{C(1, 1) - \frac{(1+T)(1-\sigma)+2\sigma}{2[(1-\sigma)+\sigma T]} C(1, w)} \right],$$

where $T \equiv t^\alpha/(1-\alpha)$ is a transformed measure of transport costs for final goods. Here, firms with productivity below some $\Theta(HH, HS)$ will conduct all activity at home, firms with intermediate productivity between $\Theta(HH, HS)$ and $\Theta(HS, SS)$ will produce intermediates

⁷We provide the algebraic expressions and more detailed arguments to support this and other claims made in this section in the appendix.

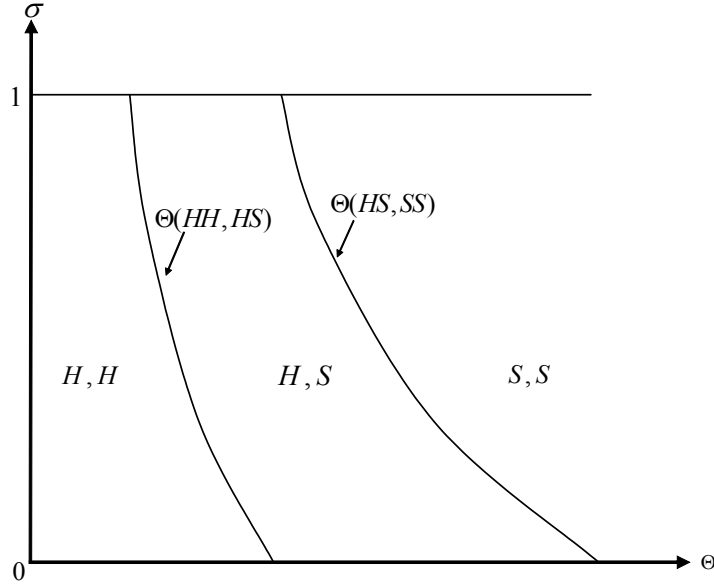


Figure 4: Optimal integration strategies when t is small and g/f is large

at home and assemble in South, and firms with productivity above $\Theta(HS, SS)$ will perform all production activities in South. The figure shows that $\Theta(HH, HS)$ is a decreasing function of σ , inasmuch as assembly in the South is more profitable when the Southern market is relatively large. Also, $\Theta(HS, SS)$ is a decreasing function of σ for reasons that are a bit more subtle. Comparing a strategy that has only assembly performed in South with a strategy that has all production in South, the latter provides lower unit costs and therefore larger volumes of output. With larger volumes, the transport cost savings from assembling in South are larger the greater is the share of South in the world market. Therefore, the break-even point between a strategy of performing only assembly in South and a strategy of performing all production activity in South will come at a lower productivity level when the market share of the South is larger.

When the fixed cost of producing intermediate goods in South is small relative to the fixed cost of foreign assembly, an integration strategy with production of intermediates in South and assembly at home will be optimal for firms with intermediate productivity levels. Figure 5 shows the optimal integration strategies for different firms when t is small and

$$\frac{g}{f} \leq \frac{C(w, w)}{C(1, 1)} \left[\frac{C(1, 1) - C(w, 1)}{\frac{2[(1-\sigma)+\sigma T]C(w, 1)}{(1+T)(1-\sigma)+2\sigma} - C(w, w)} \right].$$

As the figure shows, $\Theta(HH, SH)$ is an increasing function of σ . This reflects the fact that the amount that must be shipped to South from an assembly plant in H increases with the share of the South in the world market. But the associated shipping costs cut into the cost savings

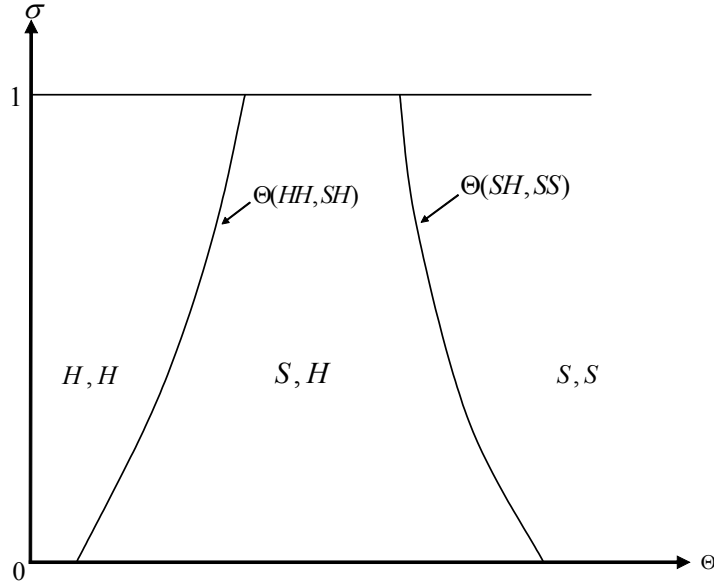


Figure 5: Optimal integration strategies when t is small and g/f is small

generated by a strategy of foreign production of intermediates, and so a higher productivity level is needed to justify the higher fixed costs of this strategy.

4.2 Moderate Transport Costs

Next we consider an industry with moderate transport costs such that

$$\frac{c(1, 1)}{c(1, w)} < t < \frac{c(w, 1)}{c(w, w)}. \quad (11)$$

When transport costs are in this range, a market in the North can be served at lower per-unit cost by exports from the South than by local assembly if and only if the intermediate goods also have been produced in the South. If instead the intermediate goods have been produced in the North, the home market can be served most cheaply via local assembly.

Again, it is never optimal for a firm with headquarters in H to produce its intermediate goods in R . Such a firm could instead produce the intermediate goods in S and achieve lower variable costs while incurring the same fixed cost. Thus, all of the integration strategies that we will consider in this section involve production of intermediates either in H or in S .

A firm that chooses to produce its intermediate goods in H will serve its home market with final goods that have been assembled there as well, in view of the left-most inequality in (11). Also, a firm that chooses to produce its intermediate goods in S will either perform all of its assembly there or else assemble all final goods at home. With intermediate goods from the South, assembly in South offers the lowest variable cost of serving any market in

view of the right-most inequality in (11). Thus, a firm that elects to bear the fixed cost of FDI in assembly will serve all markets from there. But a firm may choose to avoid the fixed cost of FDI in assembly by performing its assembly at home. We are left with six integration strategies to consider when transport costs are moderate: Southern production of intermediate goods with assembly either in H or in S ; or home production of intermediate goods with assembly in H , in H and S , in H and R , or in H, S and R .

Let us begin once again, by considering the operating profits that can be achieved by concentrating all production activities either in H or in S . By performing all activities at home, a firm avoids all fixed costs of FDI but bears a very high per-unit cost of $tc(1, 1)$ of serving the markets in R and S , and a reasonably high per-unit cost of $c(1, 1)$ of serving the home market. Nonetheless, this strategy will be attractive to firms with very low productivity, because these firms intend to produce low volumes of output. The associated operating profits are given by

$$\pi_{H,H} = (1 - \alpha)\bar{Y}\Theta \frac{[(\frac{1-\sigma}{2})(1+T) + \sigma]}{TC(1, 1)} .$$

At the other extreme, by performing all activities in South, a firm pays a high total fixed cost of $f + g$, but it attains the lowest possible per-unit cost of serving each of the markets. Operating profits then are given by

$$\pi_{S,S} = (1 - \alpha)\bar{Y}\Theta \frac{[(1 - \sigma) + \sigma T]}{TC(w, w)} - (f + g) \quad (12)$$

Such a strategy will appeal to firms with high productivity that intend to produce great volumes of output. It follows, as before, that the lowest productivity firms concentrate their activities in the home country and the highest productivity firms perform all production activities in the low-wage South.

Next consider a strategy that involves production of intermediate goods in the home country and assembly in H and in at least one other country. If assembly takes place only in H and R , the firm is engaged in horizontal FDI to conserve on shipping costs to this market. The resulting profits are⁸

$$\pi_{H,HR}(\Theta) = (1 - \alpha)\bar{Y}\Theta \frac{[(1 - \sigma)T + \sigma]}{TC(1, 1)} - f . \quad (13)$$

If assembly takes place only in H and S , the firm uses its plant in S to serve the Southern market and as an export platform for sales to R . Then operating profits are given by

$$\pi_{H,HS}(\Theta) = (1 - \alpha)\bar{Y}\Theta \left[\frac{\frac{1-\sigma}{2}}{C(1, 1)} + \frac{\frac{1-\sigma}{2} + \sigma T}{TC(1, w)} \right] - f . \quad (14)$$

⁸In this notation, the subscript on π gives the index of the country (or countries) in which the firm produces its intermediates followed by a comma and then a list of the countries in which assembly takes place.

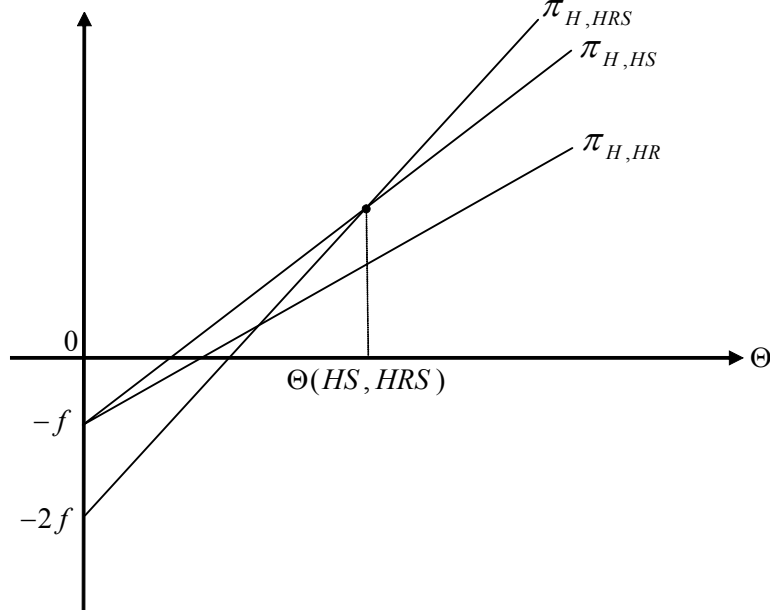


Figure 6: Assembly in multiple plants with moderate transport costs

Finally, if assembly takes place in every country, each market is served by products assembled locally, and operating profits are given by

$$\pi_{H,HRS}(\Theta) = (1 - \alpha)\bar{Y}\Theta \left[\frac{1 - \sigma}{C(1,1)} + \frac{\sigma}{C(1,w)} \right] - 2f . \quad (15)$$

Figure 6 depicts the operating profits for the integration strategies that involve assembly in more than one location. Of the three, the strategy in which the firm operates assembly plants in all three countries has the highest fixed cost and the lowest per-unit variable cost. The variable costs are low with this strategy, because the firm avoids all shipping costs. The strategy is preferred to the other two by firms with relatively high productivity. The remaining two strategies entail similar fixed costs of FDI. The figure shows a case in which a strategy of assembling in S for sales in S and R generates lower variable costs and therefore higher operating profits than a strategy of assembling in R for these markets.⁹ This case applies whenever the market share of the South is greater than $\hat{\sigma}_H$, where

$$\hat{\sigma}_H = \frac{TC(1,w) - C(1,1)}{(2T-1)C(1,1) + (T-2)C(1,w)}$$

is the critical value of σ at which it is equally profitable to assemble in H and R as it is to assemble in H and S , when intermediate goods are produced in H . If $\sigma < \hat{\sigma}_H$, then

⁹Equivalently, the firm might assemble in R for sales in R and serve the market in S with exports from H . Once the fixed cost of an assembly plant in R has been borne, the cost of exporting to S from R or H are the same.

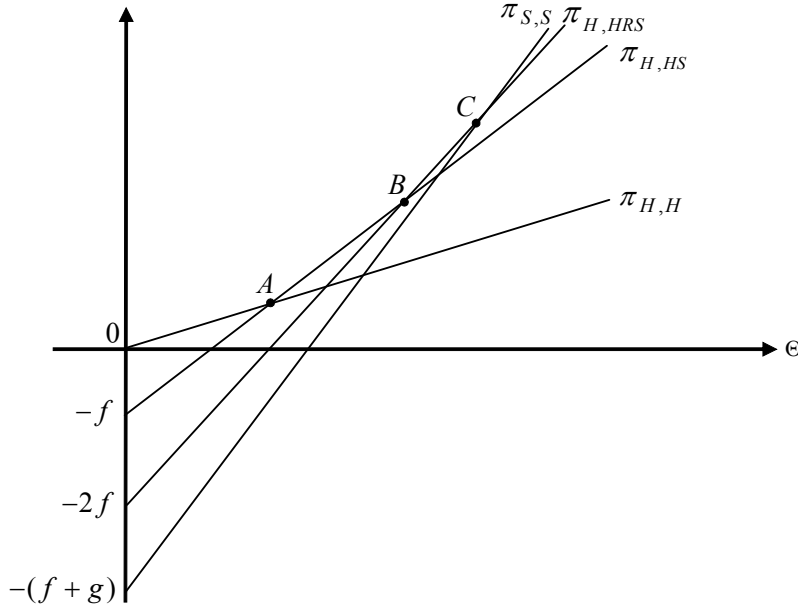


Figure 7: Moderate transport costs, $\sigma > \hat{\sigma}_H$, and g/f large

$\pi_{H,HR} > \pi_{H,HS}$ for all Θ .

If the profit line for $\pi_{H,H}$ were added to Figure 6, it would be apparent how firms that might choose to produce intermediates at home would locate their assembly operations. Those with low productivity prefer a single assembly plant at home, while those with high productivity prefer to have assembly plants in all three countries. The firms with intermediate levels of productivity prefer to have an assembly operation at home and in one other country; in the South if σ is large, and in R otherwise.

Now we consider a firm's option to produce its intermediates in the South and then assemble final goods in either H or S . If assembly takes place at home, operating profits are

$$\pi_{S,H} = (1 - \alpha)\bar{Y}\Theta \left[\frac{(1 - \sigma)(1 + T) + 2\sigma}{2TC(w, 1)} \right] - g,$$

whereas if assembly takes place in S , the profits are given in (12). Among these two strategies, firms with low productivity prefer the former and firms with high productivity prefer the latter.

A comparison of strategies involving production of intermediates at home and production of intermediates in South hinges on the relative sizes of the fixed costs of the two types of FDI. Take first the case with g/f large; i.e., the fixed cost of FDI in assembly is small relative to the fixed cost of FDI in producing intermediate goods. In drawing Figure 7 we also assume that σ is sufficiently large that $\pi_{H,HS} > \pi_{H,HR}$, and we reproduce the profit lines from Figure 6 for $\pi_{H,HS}$ and $\pi_{H,HRS}$. We show in this figure as well the profit lines for $\pi_{H,H}$ and $\pi_{S,S}$.

Since the fixed cost g of FDI in producing intermediates is large, the $\pi_{S,S}$ curve has a low intercept relative to the curves for $\pi_{H,HS}$ and $\pi_{H,HRS}$. This means that the intersection of $\pi_{S,S}$ with $\pi_{H,HRS}$ comes at a point C to the right of the intersection point B of $\pi_{H,HS}$ and $\pi_{H,HRS}$. In turn, this means that a range of firms with relatively high productivity (Θ greater than the ordinate of point B) but not very high productivity (Θ less than the ordinate of point C) prefer to produce their intermediates at home and assemble final goods in all three countries than to conduct all activity in South.¹⁰ In the figure, we have suppressed the profit line associated with $\pi_{S,H}$; with g/f large, this strategy has a high fixed cost relative to all strategies except that of conducting all production activities in S , and compared to that strategy it offers higher variable costs.

We can now describe the optimal integration strategies as a function Θ for the case illustrated, with t moderate, σ large, and g/f large. Firms with low productivity (Θ less than the ordinate of point A) perform all activities at home and export finished goods to the rest of the world. Firms with somewhat higher productivity (Θ between the ordinate of points A and B) produce their intermediate goods at home and also conduct assembly at home for local sales while operating an assembly plant in the South to serve the markets in S and R . These firms export intermediate goods to a foreign assembly plant and export final goods from S to R . Firms with still higher productivity (Θ between the ordinate of points B and C) produce all of their intermediates at home and assemble goods in each country for local sale. These firms trade intermediate goods, but not final goods. Finally, the most productive firms (Θ greater than the ordinate of point C) conduct all production activities in the low-wage South and use an assembly plant there as a platform for serving all markets.

We can use Figure 8 to discuss how the relative size of the South affects the choice of integration strategy, still for the case of g/f large. So far, we have described the optimal strategies for σ just above $\hat{\sigma}_H$. As σ grows still larger, the productivity level at which a firm that produces intermediates at home is indifferent between assembling only at home and assembling at home and in the South decreases, because an increase in the relative size of the Southern market makes assembly there relatively more profitable for any Θ . Also, an increase in σ raises the productivity level at which a firm that produces intermediates at home is indifferent between assembling in H and S and assembling in all three countries. For σ large enough, the latter strategy is dominated by one of the others, and there are no firms that assemble final goods in country R . A large σ means a relatively small market in country R , and eventually either the market there is so small that the transport cost savings do not justify the fixed cost of FDI in this country for even relatively high productivity firms, or the market in S is sufficiently large to warrant shifting all production to this location. Note that an increase in the size of the South eventually leads to not only more FDI in assembly, but also to more FDI in the production of intermediate goods. This is true even when, as here,

¹⁰If point C were to the left of point B , no firm would choose to assemble final goods in all three countries.

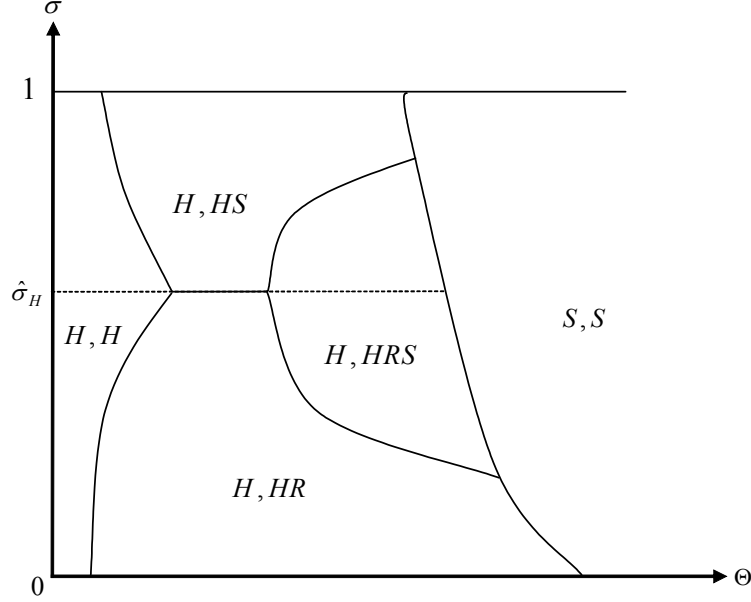


Figure 8: Optimal integration strategies with moderate transport costs and g/f large

the intermediate goods can be shipped costlessly. The two types of FDI are complementary, because when assembly takes place in the South, a larger σ spells lower average transport costs and thus greater volumes of output at a given productivity level. With more output being produced, the cost savings promised by the South's low wages justify FDI in intermediate goods at lower levels of productivity.

Now consider a Southern market share σ just below $\hat{\sigma}_H$. For all $\sigma < \hat{\sigma}_H$, a strategy with production of intermediate goods at home and assembly in H and S is dominated by one with production of intermediate goods at home and assembly in H and R . For σ slightly less than $\hat{\sigma}_H$, the choice of integration strategies is similar to those for σ just above $\hat{\sigma}_H$, except that some firms with intermediate productivity levels choose to assemble in H and R , rather than in H and S . As σ shrinks further, there are fewer firms that choose to assemble in all three countries, and for σ below some level, no firms find this to be an optimal strategy. Moreover, the smaller is the South, the greater is the productivity level needed before a firm opts to move all production activity there. The market share of the South also affects the decision of firms that manufacture intermediates at home and are choosing between assembling only in H and assembling in H and in R . As σ shrinks for given \bar{Y} , the size of the markets in H and R grow. The growth in the size of the home market has no effect on the relative profitability of these two strategies, but the growth in the size of the market in R favors the strategy with assembly there. Accordingly, the productivity level at which a firm is indifferent between assembly in H and R and assembling only in H falls as σ shrinks to zero.

Let us consider briefly an industry with moderate transport costs and g/f small. In this

case, the fixed cost of FDI for producing intermediate goods is much smaller than that for FDI in assembly. Since intermediate goods also are costless to ship, all firms except those with very low levels of productivity will find it optimal to produce their intermediate goods in the South. In particular, with g/f small, any strategy with production of intermediate goods at home and assembly in South or in multiple locations is dominated by a strategy with all production activities concentrated at home or all production activities concentrated in the South. And any strategy with production of intermediates in the South and assembly in multiple locations is dominated by one with assembly only at home or only in the South.

It follows that all firms choose one of three integration strategies: either all activity is concentrated at home, all activity is concentrated in South, or intermediates are manufactured in the South and assembled at home. It should be clear by now how firms with different productivity levels divide among these three alternatives. Those with very low productivity opt to minimize their fixed costs by conducting all production activities at home. Those with very high productivity minimize their variable costs by performing all production activities in the low-wage South. And those with intermediate levels of productivity produce their intermediate goods in the South and assemble them at home. Compared to the option of conducting all activity in the South, such firms accept the relatively high variable cost of sales to all three markets in order to avoid the relatively large fixed costs of FDI in assembly.

4.3 High Transport Costs

Finally, we consider an industry in which shipping final goods is quite costly, so that

$$t > \frac{c(w, 1)}{c(w, w)} . \quad (16)$$

In such circumstances, the lowest variable cost of serving any market is achieved by local assembly near to consumers.¹¹

Suppose first that g/f is large; i.e., the fixed cost of FDI in producing intermediate goods is large compared to the fixed cost of FDI in assembly. Then it will not be optimal for any firm to produce intermediates in a foreign subsidiary in South and conduct assembly only at home. Such a strategy is dominated either by producing the intermediates in South and conducting assembly in every country or by performing all production activities at home. The same two strategies together dominate one of producing intermediates in S and performing assembly in H and S , unless σ is close to one. Similarly, no firm will choose to assemble in H and R intermediate goods produced in S unless σ is quite small. And, as before, there exists a critical value of σ (that we have denoted by $\hat{\sigma}_H$) at which a strategy of producing intermediates at home and performing assembly in H and R yields the same operating profits

¹¹Recall that an elasticity of substitution between intermediates and assembly of no more than one ensures $c(w, 1)/c(w, w) > c(1, 1)/c(1, w)$. Therefore, when (16) is satisfied, $tc(1, w) > c(1, 1)$.

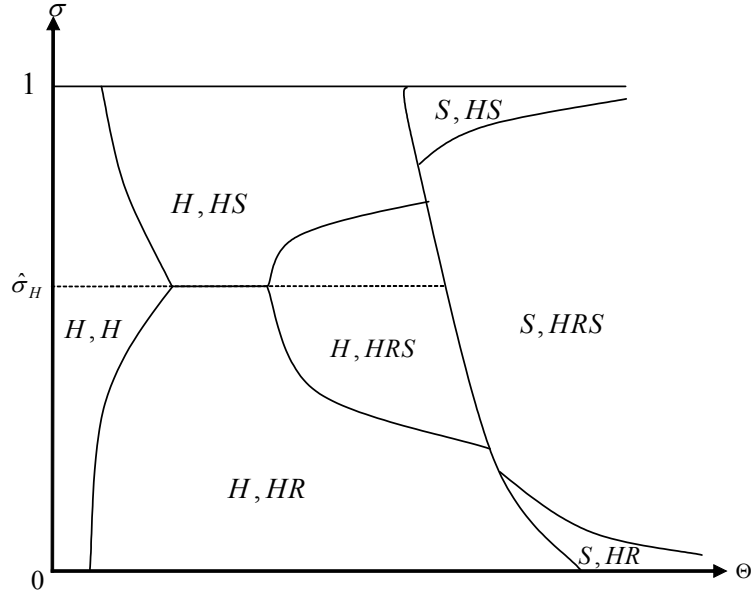


Figure 9: Optimal integration strategies with large transport costs and g/f large

as one of producing intermediates at home and performing assembly in H and S .

Consider σ slightly below $\hat{\sigma}_H$ and refer to Figure 9. As usual, firms with low productivity can minimize fixed costs by conducting all activities at home. The next lowest fixed cost is achieved by adding a single assembly plant. Since $\sigma < \hat{\sigma}_H$, every firm prefers to add such a plant in R rather than in S . By doing so, it can save on the variable costs of serving the market in R at the relatively-low fixed cost of f . This strategy will be chosen by firms that have productivity levels that are low but not quite as low as those that choose to stay entirely at home. Still more productive firms will add an assembly plant in S so as to reduce the variable costs of serving that country's market. Finally, the most productive firms are willing to pay the high fixed cost g of opening a subsidiary to produce intermediate goods in South. These firms assemble the intermediates from S in separate plants located near consumers, thereby minimizing the variable cost of serving every market.

Now let us reduce the relative size of the market in S . Since we maintain the assumption that the two Northern countries are symmetric, a fall in σ increases the size of R . Accordingly, the productivity level at which a firm is willing to open an assembly operation in R falls. Also, as σ falls, so does the profitability of operating an assembly plant in S . The productivity level at which a firm that produces its intermediates in H is indifferent between performing assembly only in H and R and doing so also in S rises, until eventually a value of σ is reached such that no firm that produces intermediates in H assembles final goods in the South. When σ is quite small, even a firm that produces intermediates in S will not opt to assemble there unless θ is very large. Rather, for σ small, there exists a range of quite high productivity

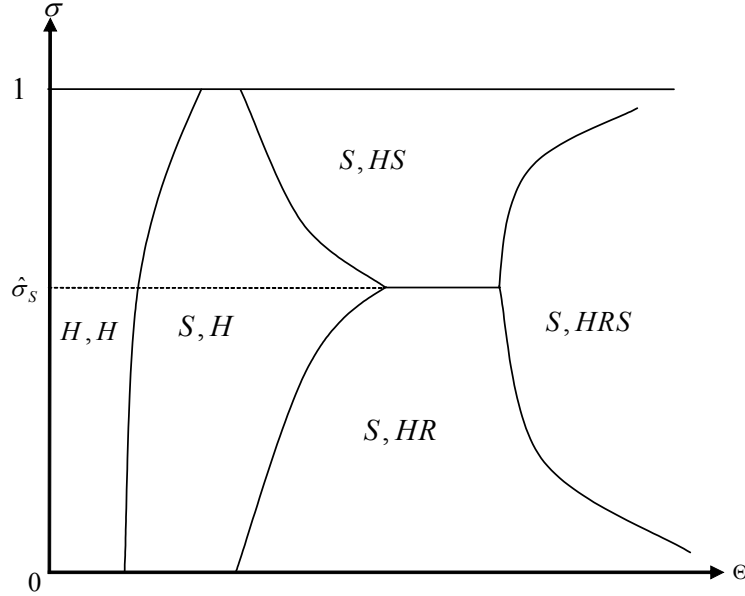


Figure 10: Optimal integration strategies with large transport costs and g/f small

levels at which firms produce their intermediates in S and assemble them only in the two Northern countries. These firms serve the small Southern market with exports from one or the other of their Northern plants.

The portion of Figure 9 that applies for $\sigma > \hat{\sigma}_H$ can be understood similarly. Since we believe that most industries of interest have relatively large markets in the North, we will not go through the analysis in detail. Suffice it to say that, with g/f large, only the most productive firms find it worthwhile to produce intermediates in the South. And only the least productive firms do not want to serve at least the larger of the two foreign markets—if not both—with goods produced in an assembly plant located near consumers. The lowest variable cost is achieved by producing intermediates in the South and assembling them in all three countries. Thus, a highly productive firm will opt for such a strategy despite the high fixed costs of $g + 2f$ that are entailed.

It remains to discuss an industry with high transport costs and g/f small. The optimal strategies for firms in such an industry are shown in Figure 10. Since the fixed cost of FDI for producing intermediates is relatively low, only the least productive firms opt to manufacture their intermediate goods at home. Those with productivity just above these low levels produce intermediates in the South and export them from S to an assembly plant in their home country. Such a strategy involves a relatively low fixed cost of g , but entails relatively high variable costs of serving the markets in R and S due to the high transport costs. Accordingly, more productive firms will opt to open an assembly plant abroad; in R

if $\sigma < \hat{\sigma}_S$ and in S otherwise.¹² Finally, the most productive firms operate assembly plants in both R and S . These firms export intermediate goods from S to H and to R , but do not trade any of their final products.

The slopes of the various boundary lines in the figure are easy to understand in the light of the previous discussion. For example, the boundary between (S, H) and (S, HS) is downward sloping in the region with $\sigma > \hat{\sigma}_S$, because the profitability of an assembly plant in S increases with an increase in South's market share. Similarly, for $\sigma < \hat{\sigma}_S$, the boundary between (S, H) and (S, HR) slopes upward, because an increase in σ corresponds to a decrease in the relative size of R and so firms require a higher level of productivity to justify placing an assembly plant there. The slopes of the boundaries between (S, HS) and (S, HRS) and between (S, HR) and (S, HRS) have analogous explanations.

4.4 Transport Costs and FDI When the South is Small

In this section, we highlight the relationship between the cost of transporting final goods and the integration strategies chosen by multinational firms in situations where the market in the South is small. Specifically, we focus on the case where $\sigma = 0$ and review systematically how changes in t affect the relative prevalence of different organizational forms.

We consider first the case in which g/f is large; i.e., the fixed cost of FDI in intermediate production is large relative to the fixed cost of FDI in assembly. When transport costs are small and $\sigma = 0$, the optimal integration strategies for firms with different productivity levels are shown along the horizontal axis of Figure 4. We see that firms with low productivity conduct all activity at home, firms with intermediate productivity produce intermediate goods in H and assemble final goods in S , while firms with high productivity conduct all production activity in the South. As the cost of transport rises but still remains in the range where $t < t_m \equiv c(1, 1)/c(1, w)$, the boundary $\Theta(HH, HS)$ at which the first two of these strategies are equally profitable shifts to the right, as does the boundary $\Theta(HS, SS)$ at which the last two strategies are equally profitable. An increase in transport cost favors assembly at home relative to assembly in the South, because the Southern market is small and shipments from South to Home become more expensive as t rises. Thus, equal profitability occurs at a higher productivity level when t is a bit larger compared to when t is close to one. The rightward shift of $\Theta(HS, SS)$ reflects that output of final goods is larger when intermediate goods are produced in the South as compared to when they are produced in the higher-cost North. Therefore, an increase in transport costs for final goods has a larger impact on the profitability of a firm that conducts all of its production activities in the South than it does on one that only performs assembly there.

We use $\gamma_{i,j}$ to denote the fraction of firms that produce intermediate goods in country

¹²We use $\hat{\sigma}_S$ to denote the critical value of σ at which it is equally profitable to assemble in H and R as it is to assemble in H and S , when intermediate goods are produced in the South.

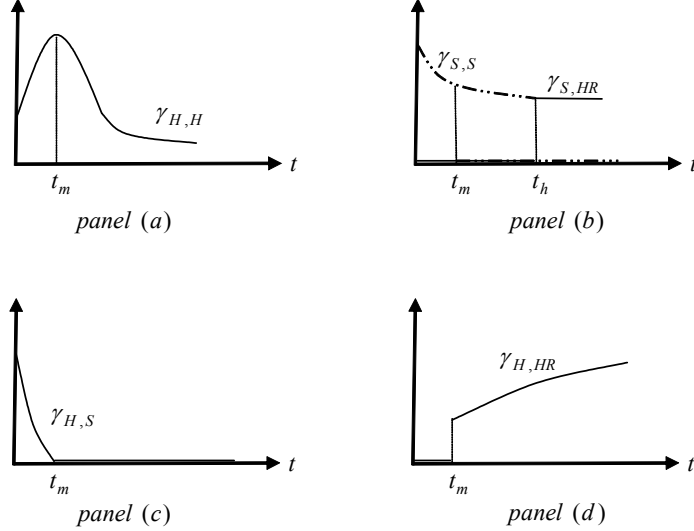


Figure 11: Relative prevalence of organizational forms when $\sigma = 0$ and g/f is large

i and assemble final goods in the set of countries $\{j\}$. Evidently, a rise in t increases $\gamma_{H,H}$ for $t < t_m$, as shown in panel (a) of Figure 11. Also, $\gamma_{S,S}$ falls, as shown by the broken curve in panel (b). The fraction of firms that produce intermediate goods in the North and assemble in the South can rise or fall, since this organizational form gains at the expense of integrated production in the South but loses at the expense of integrated production in the North. However, it can readily be shown that $\gamma_{H,S}$ must fall as t rises when the size distribution of productivity levels is characterized by a Pareto distribution.¹³ This is shown in panel (c).

Next suppose that t rises above t_m . For $t \geq t_m$, a firm that produces its intermediate goods in H can earn higher profits by performing assembly in H and R than by performing assembly in S . The integration strategies used by firms with different productivity levels are shown along the horizontal axis of Figure 8. Firms with low productivity conduct all activity at home, firms with intermediate productivity produce intermediate goods at home and assemble in H and R , and firms with high productivity conduct all production activity in the South. Further increases in t for $t_m < t \leq t_h \equiv c(w, 1)/c(w, w)$ cause a contraction in the fraction of firms with all production activities in H or S , and an expansion in the fraction of firms that assemble in both H and R . Thus, $\gamma_{H,H}$ and $\gamma_{S,S}$ fall while $\gamma_{H,HR}$ rises with

¹³With a Pareto distribution of θ , the fraction of firms that have productivity less than θ is given by $G(\theta) = 1 - (\tilde{b}/\theta)^{\tilde{k}}$ for $\theta \geq \tilde{b} > 0$ and $\tilde{k} > 1$. We show in the appendix that with this distribution the share $\gamma_{H,S}$ is a declining function of transport costs when g/f is large.

The Pareto distribution is commonly used to describe the size distribution of productivity levels. Axtell (2001) provides evidence that such a distribution fits well the data on the distribution of sales by U.S. firms. Helpman et al. (2003) show that a Pareto distribution of firm sizes emerges from a Pareto distribution of productivity measures.

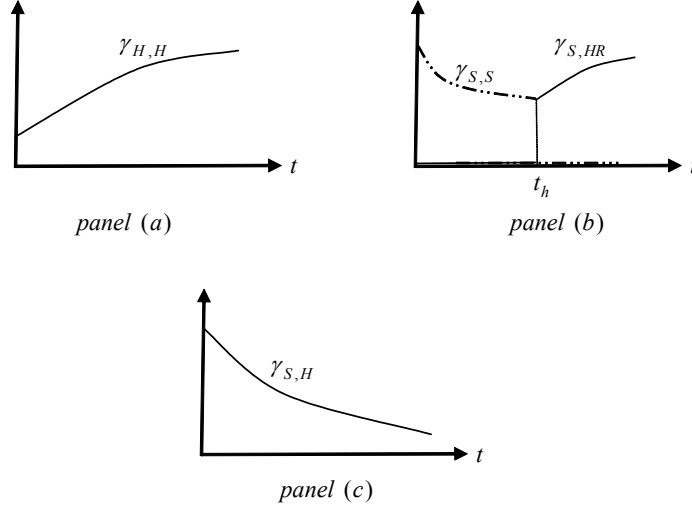


Figure 12: Relative prevalence of organizational forms when $\sigma = 0$ and g/f is small

increases in transport costs in this range, as shown in panels (a), (b) and (d) of Figure 8.

Finally, for high transport costs such that $t > t_h$, the optimal integration strategies are shown in Figure 9. For $t \geq t_h$, no assembly is performed in the South. Now, even the firms with high productivity that produce intermediates in S find it more profitable to assemble their final goods close to consumers. The low productivity firms perform all activities at home, the firms with intermediate productivity manufacture intermediate goods at home and assemble them in H and R , and the high productivity firms produce intermediate goods in the South and assemble in H and R . Further increases in transport costs continue to shrink the range of productivities for which it is most profitable to have a fully integrated facility in the home country and serve R by exports. However, the productivity level at which it is equally profitable to produce intermediates in H or S , with assembly in each case in H and R , is unaffected by changes in t , inasmuch as neither of these strategies involves any trade in final goods. Figure 8 shows $\gamma_{H,H}$ falling, $\gamma_{S,HR}$ flat, and $\gamma_{H,HR}$ rising for $t \geq t_h$.

Overall, increases in transport costs for final goods cause production activities to shift away from the South. Assembly operations in the South are replaced by assembly operations closer to consumers. And production of intermediate goods in the South contracts in favor of production at home, inasmuch as the strategies that involve intermediate production in the South also involve high volumes of output and so are especially vulnerable to increases in the cost of transporting final goods.

We turn briefly to the case of g/f small. When the fixed cost of FDI in assembly is relatively large, the same three integration strategies are observed in equilibrium for all $t < t_h$; firms with low productivity conduct all activities at home, firms with intermediate productivity produce intermediate goods in a subsidiary in the South but perform assembly

at home, and firms with high productivity perform all production activities in the South. An increase in t for $t < t_h$ causes both $\Theta(HH, SH)$ and $\Theta(SH, SS)$ to shift to the right. The range of firms that performs all activity at home expands, because a firm that manufactures its intermediate goods in the South will choose to produce more output than one with similarly productivity that manufactures these goods at home, so the rise in the cost of transporting final goods from H to R will impact more strongly the multinational firms. A strategy of integrated production in the South becomes less profitable, because this strategy involves the transport of final goods to markets in both H and R . Thus, as t grows, $\gamma_{H,H}$ rises and $\gamma_{S,S}$ falls, as shown in panels (a) and (b) of Figure 12. The fraction of firms that produces intermediate goods in the South and assembles them at home may rise or fall with an arbitrary distribution of productivity levels; but it must decline if θ has a Pareto distribution, as shown in panel (c) of Figure 12.¹⁴

For $t > t_h$, any strategy with assembly in the South is dominated by one with assembly in H and R . Thus, for transport costs in this range, firms with low productivity concentrate all activity at home, firms with intermediate productivity produce intermediates in the South and assemble at home, and firms with high productivity manufacture intermediate goods in the South and conduct assembly near to their markets in H and R . A further increase in t expands the range of productivity levels for which firms concentrate all activities at home, as well as that for which assembly in H and R is optimal; thus, $\gamma_{H,H}$ and $\gamma_{S,HR}$ increase with t , while $\gamma_{S,H}$ declines, as shown in Figure 12. As t rises, the fraction of multinational firms falls monotonically, as production of intermediate goods in South gives way to production of these goods at home, and assembly in the South gives way to assembly at home and ultimately to assembly in both Northern markets.

5 Transport Costs for Intermediate Goods

Up until now, we have assumed that intermediate goods can be moved costlessly to any place of assembly. This simplifying assumption allowed us to examine how variations in the cost of transporting final goods, in relative market size, and in the relative fixed costs of FDI in different activities affect firms' decisions about global integration. We have seen that, with variations in productivity in an industry, rich patterns of trade and FDI are possible.

In this section, we introduce a cost of trading intermediate inputs. To avoid a detailed taxonomy, however, we explore only cases in which the cost of transporting intermediates goods is high. Such costs give firms an incentive to locate the production of intermediate goods near to where they intend to perform their assembly activities. We also assume that the South is negligible in size ($\sigma = 0$), so that firms are not motivated to locate their assembly operations in S in order to serve the Southern market. Rather, if a firm opens an assembly

¹⁴See the appendix for a derivation of this result.

plant in the South, it is because it wishes to use such a plant as an export platform.

Recall that τ captures the cost of shipping intermediate goods; $\tau > 1$ units of the good must be shipped from a production facility in some country to deliver one unit of the good to an assembly plant in another country. The sense in which we assume that τ is large is that

$$wc(\tau, 1) > c(1, 1) . \tag{17}$$

This restriction implies that $\tau w > 1$; i.e., it is more costly to produce the intermediate good in South and ship it to the North than it is to manufacture the good in East or West.¹⁵ Also, it implies that $c(\tau, w) > c(1, 1)$; i.e., a final good assembled in S with intermediates imported from the North has a greater per-unit cost than one produced and assembled entirely in a single Northern country.¹⁶

With high costs of transporting intermediate goods and a small market in the South, many integration strategies can be ruled out. First, no firm with its headquarters in H will have a sole assembly plant in country R . Any such strategy is dominated by an alternative one with all production activities undertaken at home.¹⁷ Second, no firm will maintain assembly plants in all three countries, inasmuch as the plant in the South would then serve only the negligible Southern market and thus would not cover its fixed costs. Third, if a firm has assembly plants in H and R , it will not produce intermediates only in R . Instead, it could produce intermediates only in H , achieve the same total variable costs and conserve on fixed costs. Fourth, a firm with assembly plants in H and R will not produce any intermediate goods in S , because delivery from S involves higher costs than local production near an assembly plant and entails fixed costs that are at least as large. Taken together, these observations imply that any firm with an assembly operation in R will also have one in H , and that such a firm will produce intermediate goods either only in H or in H and R . The resulting profits are $\pi_{H,HR}$ or $\pi_{HR,HR}$, as the case may be.

Four other strategies seem viable at this point. First, a firm may assemble only in S . But then it will produce its intermediates in S as well, because a strategy with intermediates produced in H or R and assembled in S is dominated by one with all production activities in H , inasmuch as (17) implies $c(\tau, w) > c(1, 1)$. Second, a firm may assemble only in H . Then it will produce its intermediate goods in H as well, to conserve on both shipping costs and fixed costs. Finally, a firm may operate assembly plants in H and S , the former to serve the home market and the latter to serve the market in R . If so, it will produce intermediates

¹⁵Note that $wc(\tau, 1) = c(\tau w, w)$. Then $c(\tau w, w) > c(1, 1)$ and $w < 1$ implies $\tau w > 1$.

¹⁶Note that $c(\tau, w) > c(\tau w, w) = wc(\tau, 1)$.

¹⁷If the sole assembly plant is in R , there would be no reason to have multiple plants producing intermediate goods. Then a strategy of producing intermediates in S and assembling in R is dominated by one with all production activities in R , because $\tau w > 1$. And a strategy with all production in R or one with production of intermediates in H and assembly in R is dominated by one with all production in H ; the latter strategy yields variable costs that are no higher (in view of the symmetry of the two countries) and conserves on fixed costs of FDI.

either in S only, or in H and S .¹⁸ The four strategies that do not involve assembly in R yield potential profits of $\pi_{S,S}$, $\pi_{H,H}$, $\pi_{S,HS}$, and $\pi_{HS,HS}$, respectively.

The optimal choices among the remaining strategies will depend on the sizes of the fixed costs of FDI in intermediates and assembly, the cost of transporting final goods, and the productivity of the firm. Consider first an industry with a low cost of transporting final goods such that

$$t < \frac{1}{w} . \quad (18)$$

This implies $tc(w, w) < c(1, 1)$ or that the variable cost of serving any Northern market from an integrated production and assembly plant in the South is less than the cost of serving the same market from an assembly plant in the North.¹⁹ Thus, $\pi_{S,S}(\theta) > \pi_{HS,HS}(\theta) > \pi_{HR,HR}(\theta)$ and $\pi_{S,S}(\theta) > \pi_{S,HS}(\theta)$ for all θ . Moreover, (17) and (18) together imply $tc(1, 1) < c(\tau, 1)$, so that $\pi_{H,H}(\theta) > \pi_{H,HR}(\theta)$. In other words, the only viable integration strategies when t is small are ones with all production activities concentrated either in South or at home. Relatively productive firms will prefer the former strategy with its lower variable costs, while less productive firms prefer the latter strategy with its lower fixed costs. Evidently, the high cost of transporting intermediate goods dictates the optimal strategy for integration, which must involve production of intermediate goods and assembly in the same location.

Next consider intermediate costs of shipping final goods such that

$$\frac{1}{w} < t < \frac{c(\tau, 1)}{c(1, 1)} . \quad (19)$$

With $t > 1/w$, a strategy of integrated production and assembly in both H and R offers higher profits to any firm than one of concentrating all production in the South, operating integrated production and assembly operations in both H and S , and producing intermediates in S with assembly in H and S . This is so, because all of these strategies entail the same fixed cost $f + g$ and with relatively costly transport of intermediate and final goods it is better to assemble output near to consumers than to ship output from the South.

The right-most inequality in (19) implies that the variable cost of supplying the market in R with goods produced entirely in H is less than the variable cost of serving that market with goods assembled there with intermediates imported from H . This in turn implies that $\pi_{H,H}(\theta) > \pi_{H,HR}(\theta)$, since concentrating production in H also conserves on fixed costs. Thus, the only viable strategies when t satisfies (19) are ones with an integrated production and assembly operation only in H or with such integrated operations both in H and in R .

¹⁸If assembly is performed in H and S , a strategy with production of intermediates only in R or in R and S is dominated by one with production only in H or in H and S , respectively. A strategy with intermediates produced in H and R is dominated by one with intermediates produced in H and S . And no firm will choose to produce intermediates only in H and assemble in H or S , because with intermediates emanating only from H the firm could earn higher profits by serving the market in R from an assembly plant located there.

¹⁹The variable cost of serving a Northern market from an assembly plant located in that market is at least $c(1, 1)$.

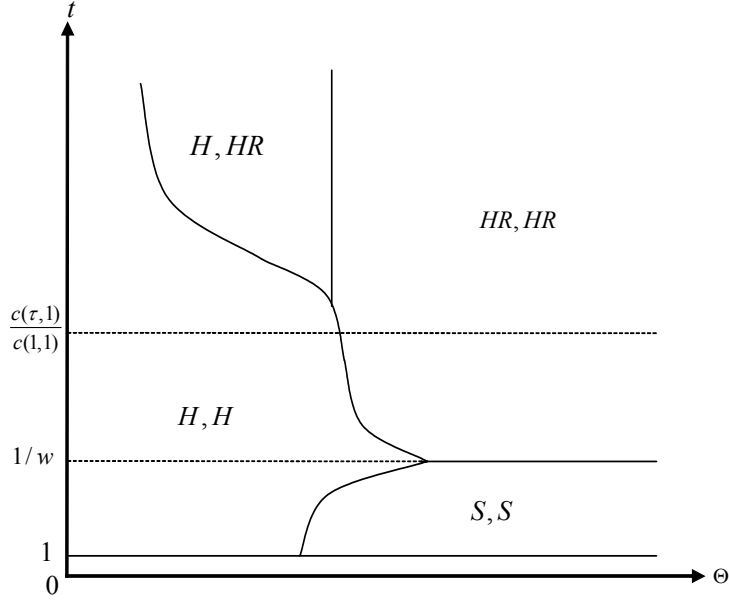


Figure 13: Optimal integration strategies when τ is large and $\sigma = 0$

The former strategy is preferred by firms with relatively low productivity, the latter by firms with relatively high productivity.

Finally, for large costs of transporting final goods that satisfy

$$t > \frac{c(\tau, 1)}{c(1, 1)} \quad (20)$$

it may be optimal for some firms to assemble goods in R with intermediate inputs imported from H . As before, $t > 1/w$ eliminates from consideration all strategies with assembly operations in S . So, the three viable strategies for t satisfying (20) yield operating profits of $\pi_{H,H}$, $\pi_{H,HR}$, and $\pi_{HR,HR}$. The first of these minimizes the fixed costs while the last minimizes the variable costs of serving both markets. A strategy of assembly in R with intermediates produced in H conserves on fixed costs relative to one with intermediates produced in H and R , but it generates a higher variable cost of serving the market in R . Accordingly, firms with low productivity will conduct all activity in H , firms with intermediate productivity will assemble final goods in H and R but produce intermediate goods only in H , and firms with high productivity will operate integrated production and assembly facilities in both H and R .

Figure 13 summarizes the discussion in this section and also indicates how the choices of a firm with a given productivity level are affected by the size of the transport cost for final goods. For $t < 1/w$, the greater is the transport cost the higher must be a firm's productivity before it will choose to shift operations to the South. For t in the intermediate range, the

greater are transport costs the lower is the productivity level at which a firm will open an integrated production facility in R . And finally, for t large, the greater are transport costs, the lower is the productivity level at which a firm is indifferent between maintaining an assembly plant in R and not in H . The cost of shipping final goods has no bearing on firms' decision whether to produce intermediate goods in H only, or in H and R , since this decision applies to firms that are not trading final goods in any case.

6 Concluding Remarks

In this paper, we have examined the joint determination of international trade and foreign direct investment in a setting in which firms may choose among a rich array of integration strategies. In our analysis, firms that are headquartered in a Northern country supply differentiated final goods to two national markets in the North and one in the South. Each such firm must produce an intermediate input and conduct assembly activities in order to generate a final product. The firms may produce intermediate goods in their home country, in the other Northern country, or in the South. Similarly, assembly may take place in any of the three locations. And firms may choose to maintain plants for either or both stages of production in multiple locations. Accordingly, there are many possible choices of organizational forms available to firms. Each such choice has implications for the pattern of trade in intermediate and final goods.

We characterized industries by the size of the fixed costs of maintaining a foreign subsidiary for production of intermediate goods and for assembly, the cost of transporting intermediate and final goods internationally, and the fraction of the consumer demand that resides in the low-wage South. For each industry, we derived the equilibrium organizational forms of heterogeneous firms that differ in their productivity.

In an industry in which transportation of intermediate and final goods is costless, the relative size of fixed costs for foreign investment in intermediate goods and assembly determines the set of organizational firms that are observed in equilibrium. Here, the relative sizes of the markets have no bearing on the equilibrium choices, and there is no intra-industry FDI. Firms with low productivity choose an integration strategy that minimize the fixed cost of operation, whereas firms with high productivity seek to minimize variable costs of serving the various markets.

When final goods are costly to transport, relative country size plays a role in determining the viable multinational strategies. Generally, the larger is the consumer market in the South, the greater is the fraction of firms that maintain subsidiaries there, not only for performing assembly but also for producing intermediate goods. Also, the higher are transport costs for final goods, the greater is the fraction of firms that performs assembly in two or more locations. Finally, costly transport of intermediate goods can make it attractive for a firm to

produce intermediate goods in multiple locations. In all cases, the attractiveness of alternative locations for each stage of production depends upon the choices contemplated for the other.

One limitation of our analysis in this paper is that we take the boundaries of the firm as given. That is, we have simply assumed that firms must produce their own intermediate goods and perform assembly in-house. In other recent work (Grossman and Helpman, 2002, 2003, 2004) two of us have studied how contracting problems interact with factor-price differentials and transport costs to determine which activities are outsourced and which performed within a firms' corporate boundaries. In those papers, the range of strategies open to the multinational firm was substantially narrower than here. Ultimately, we would like a theory that simultaneously explains the make-or-buy decision and the organization of the multinational firm. Such a theory could help explain the broad range of corporate strategies that are found in the firm-level data.

Appendix

DERIVATIONS FOR SECTION 3

From (4) and (5) we find that $\pi_{H,H} = \pi_{S,S}$ for the productivity level

$$\Theta(HH, SS) = \frac{f+g}{(1-\alpha)\bar{Y}} \cdot \frac{C(w,w)C(1,1)}{C(1,1) - C(w,w)}.$$

Then

$$\pi_{H,H}(\Theta(HH, SS)) = (f+g) \cdot \frac{C(w,w)}{C(1,1) - C(w,w)}$$

and

$$\pi_{H,S}(\Theta(HH, SS)) = (f+g) \cdot \frac{C(1,1)}{C(1,w)} \frac{C(w,w)}{C(1,1) - C(w,w)} - f,$$

as well as

$$\pi_{S,H}(\Theta(HH, SS)) = (f+g) \cdot \frac{C(1,1)}{C(w,1)} \frac{C(w,w)}{C(1,1) - C(w,w)} - g$$

from (6) and (8). It follows that $\pi_{H,S}(\Theta(HH, SS)) > \pi_{H,H}(\Theta(HH, SS))$ if and only if (7) holds, and $\pi_{S,H}(\Theta(HH, SS)) > \pi_{H,H}(\Theta(HH, SS))$ if and only if (9) holds.

From $\pi_{H,H} = \pi_{H,S}$ and equations (4) and (6) we have

$$\Theta(HH, HS) = \frac{f}{(1-\alpha)\bar{Y}} \cdot \frac{C(1,w)C(1,1)}{C(1,1) - C(1,w)}.$$

We can derive in similar fashion

$$\begin{aligned} \Theta(HH, SH) &= \frac{g}{(1-\alpha)\bar{Y}} \cdot \frac{C(w,1)C(1,1)}{C(1,1) - C(w,1)}, \\ \Theta(SH, SS) &= \frac{f}{(1-\alpha)\bar{Y}} \cdot \frac{C(w,1)C(w,w)}{C(w,1) - C(w,w)}, \\ \Theta(HS, SS) &= \frac{g}{(1-\alpha)\bar{Y}} \cdot \frac{C(1,w)C(w,w)}{C(1,w) - C(w,w)}. \end{aligned}$$

DERIVATIONS FOR SECTION 4.1

The profit functions $\pi_{H,H}$ and $\pi_{S,S}$ now become

$$\pi_{H,H} = \frac{(1-\alpha)\Theta}{TC(1,1)} [Y^N(1+T) + Y^S] = (1-\alpha)\bar{Y}\Theta \frac{[\frac{1-\sigma}{2}(1+T) + \sigma]}{TC(1,1)}$$

and

$$\pi_{S,S} = \frac{(1-\alpha)\Theta}{TC(w,w)} [2Y^N + TY^S] - (f+g) = (1-\alpha)\bar{Y}\Theta \frac{[1-\sigma + T\sigma]}{TC(w,w)} - (f+g).$$

Equating them yields

$$\begin{aligned}\Theta(HH, SS) &= \frac{f+g}{(1-\alpha)\bar{Y}} \cdot \left[\frac{1-\sigma+T\sigma}{TC(w, w)} - \frac{\frac{1-\sigma}{2}(1+T)+\sigma}{TC(1, 1)} \right]^{-1} = \\ &= \frac{f+g}{(1-\alpha)\bar{Y}} \cdot C(w, w)C(1, 1) / \\ &/ \left[\frac{1}{T} \left\{ C(1, 1) - \frac{C(w, w)}{2} - \sigma \left[C(1, 1) + \frac{C(w, w)}{2} \right] \right\} + \sigma \left[C(1, 1) + \frac{C(w, w)}{2} \right] - \frac{C(w, w)}{2} \right].\end{aligned}$$

This expression is increasing in $1/T$ as long as

$$\sigma < \frac{C(1, 1) - \frac{C(w, w)}{2}}{C(1, 1) + \frac{C(w, w)}{2}} = 1 - \frac{C(w, w)}{C(1, 1) + \frac{C(w, w)}{2}}.$$

Since $C(w, w) < C(1, 1)$, this inequality is satisfied for $\sigma < 1/3$, that is, when the typical market in the North is larger than in the South. Furthermore, $\Theta(HH, SS)$ is decreasing in σ because $1/T < 1$.

Now

$$\pi_{H,H}(\Theta(HH, SS)) = \frac{(f+g)}{TC(1, 1)} \cdot \left[\frac{1-\sigma+T\sigma}{TC(w, w)} - \frac{\frac{1-\sigma}{2}(1+T)+\sigma}{TC(1, 1)} \right]^{-1} \cdot \left[\frac{1-\sigma}{2}(1+T) + \sigma \right]$$

and

$$\pi_{H,S}(\Theta(HH, SS)) = \frac{f+g}{TC(1, w)} \cdot \left[\frac{1-\sigma+T\sigma}{TC(w, w)} - \frac{\frac{1-\sigma}{2}(1+T)+\sigma}{TC(1, 1)} \right]^{-1} [1 - \sigma + \sigma T] - f.$$

Therefore $\pi_{H,S}(\Theta(HH, SS)) > \pi_{H,H}(\Theta(HH, SS))$ if and only if

$$\frac{C(1, 1)}{C(w, w)} \left[\frac{C(1, w) - C(w, w)}{C(1, 1) - \frac{\frac{1-\sigma}{2}(1+T)+\sigma}{1-\sigma+T}C(1, w)} \right] < \frac{g}{f}.$$

Similarly,

$$\pi_{S,H}(\Theta(HH, SS)) = \frac{f+g}{TC(w, 1)} \cdot \left[\frac{1-\sigma+T\sigma}{TC(w, w)} - \frac{\frac{1-\sigma}{2}(1+T)+\sigma}{TC(1, 1)} \right]^{-1} \left[\frac{1-\sigma}{2}(1+T) + \sigma \right] - g$$

and $\pi_{S,H}(\Theta(HH, SS)) > \pi_{H,H}(\Theta(HH, SS))$ if and only if

$$\frac{g}{f} < \frac{C(w, w)}{C(1, 1)} \left[\frac{C(1, 1) - C(w, 1)}{C(w, 1) \frac{1-\sigma+T\sigma}{\frac{1-\sigma}{2}(1+T)+\sigma} - C(w, w)} \right].$$

We can derive $\Theta(HH, HS)$ from $\pi_{H,H} = \pi_{H,S}$, which yields

$$\Theta(HH, HS) = \frac{f}{(1-\alpha)\bar{Y}} \cdot \left[\frac{1-\sigma+\sigma T}{TC(1,w)} - \frac{\frac{1-\sigma}{2}(1+T)+\sigma}{TC(1,1)} \right]^{-1}.$$

The expression for σ in the last term is

$$\sigma = \frac{T-1}{TC(1,w)} + \frac{T-1}{2TC(1,1)},$$

implying that $\Theta(HH, HS)$ is decreasing in σ . Similarly, we find that $\Theta(HS, SS)$ is

$$\Theta(HS, SS) = \frac{g}{(1-\alpha)\bar{Y}} \cdot \frac{1}{1-\sigma+T\sigma} \cdot \left[\frac{1}{TC(w,w)} - \frac{1}{TC(1,w)} \right]^{-1},$$

which is decreasing in σ .

DERIVATIONS FOR SECTION 4.2

Equating $\pi_{H,HR}$ and $\pi_{H,HS}$ we obtain

$$(1-\alpha)\bar{Y}\Theta \frac{[(1-\sigma)T+\sigma]}{TC(1,1)} - f = (1-\alpha)\bar{Y}\Theta \left[\frac{\frac{1-\sigma}{2}}{C(1,1)} + \frac{[\frac{1-\sigma}{2}+T\sigma]}{TC(1,w)} \right] - f,$$

or equivalently,

$$\hat{\sigma}_H = \frac{TC(1,w) - C(1,1)}{(2T-1)C(1,1) + (T-2)C(1,w)}.$$

Similarly, equating $\pi_{S,HR}$ and $\pi_{S,HS}$ yields

$$(1-\alpha)\bar{Y}\Theta \frac{[(1-\sigma)T+\sigma]}{TC(w,1)} - (f+g) = (1-\alpha)\bar{Y}\Theta \left[\frac{\frac{1-\sigma}{2}}{C(w,1)} + \frac{[\frac{1-\sigma}{2}+T\sigma]}{TC(w,w)} \right] - (f+g),$$

or

$$\hat{\sigma}_S = \frac{C(w,1) - TC(w,w)}{(2-T)C(w,w) + (1-2T)C(w,1)}.$$

DERIVATIONS FOR SECTION 4.4

When the productivity draws θ have a Pareto distribution, the distribution of productivity among the population of firms who remain in the market is also Pareto, and so is the distribution of Θ among these firms. Let the distribution of Θ be $1 - (b/\Theta)^k$, where $\Theta \geq b$

and $k > 1$. Then for the case in which g/f is large,

$$\begin{aligned}\gamma_{H,S} &= b^k \left[\frac{1}{\Theta(HH, HS)^k} - \frac{1}{\Theta(HS, SS)^k} \right] \\ &= \left[\frac{b(1-\alpha)\bar{Y}}{T} \right]^k \left[\frac{1}{f^k} \cdot \left[\frac{1}{C(1,w)} - \frac{1+T}{2C(1,1)} \right]^k - \frac{1}{g^k} \cdot \left[\frac{1}{C(w,w)} - \frac{1}{C(1,w)} \right]^k \right].\end{aligned}$$

This expression is decreasing in T .

When g/f is small, we have

$$\begin{aligned}\gamma_{S,H} &= b^k \left[\frac{1}{\Theta(HH, SH)^k} - \frac{1}{\Theta(SH, SS)^k} \right] \\ &= [b(1-\alpha)\bar{Y}]^k \left[\frac{1}{g^k} \left(\frac{1}{2T} + \frac{1}{2} \right)^k \cdot \left[\frac{1}{C(w,1)} - \frac{1}{C(1,1)} \right]^k - \frac{1}{f^k} \cdot \left[\frac{1}{TC(w,w)} - \frac{\frac{1}{2T} + \frac{1}{2}}{C(w,1)} \right]^k \right].\end{aligned}$$

The derivative of this expression with respect to T is

$$\begin{aligned}\frac{\partial \gamma_{S,H}}{\partial T} &= \frac{-[b(1-\alpha)\bar{Y}]^k}{g^k} k \left(\frac{1}{2T} + \frac{1}{2} \right)^{k-1} \frac{1}{2T^2} \cdot \left[\frac{1}{C(w,1)} - \frac{1}{C(1,1)} \right]^k \\ &\quad - \frac{[b(1-\alpha)\bar{Y}]^k}{f^k} \cdot k \left[\frac{1}{TC(w,w)} - \frac{\frac{1}{2T} + \frac{1}{2}}{C(w,1)} \right]^{k-1} \left[\frac{-1}{T^2 C(w,w)} + \frac{1}{2T^2 C(w,1)} \right] \\ &= [b(1-\alpha)\bar{Y}]^k \frac{k}{T^2} \frac{1}{f^k} \cdot \left[\frac{1}{TC(w,w)} - \frac{1+T}{2TC(w,1)} \right]^{k-1} \left[\frac{1}{C(w,w)} - \frac{1}{2C(w,1)} \right] \\ &\quad - [b(1-\alpha)\bar{Y}]^k \frac{k}{T^2} \frac{1}{g^k} \left(\frac{T+1}{2T} \right)^{k-1} \frac{1}{2} \left[\frac{1}{C(w,1)} - \frac{1}{C(1,1)} \right]^k,\end{aligned}$$

which is negative as long as

$$\begin{aligned}\frac{1}{g^k} \left(\frac{T+1}{2T} \right)^{k-1} \frac{1}{2} \left[\frac{1}{C(w,1)} - \frac{1}{C(1,1)} \right]^k \\ > \frac{1}{f^k} \cdot \left[\frac{1}{TC(w,w)} - \frac{1+T}{2TC(w,1)} \right]^{k-1} \left[\frac{1}{C(w,w)} - \frac{1}{2C(w,1)} \right],\end{aligned}$$

or equivalently,

$$\frac{\frac{1}{2^{1/k}} \left(\frac{T+1}{2T} \right)^{\frac{k-1}{k}} \left[\frac{1}{C(w,1)} - \frac{1}{C(1,1)} \right]}{\left[\frac{1}{TC(w,w)} - \frac{1+T}{2TC(w,1)} \right]^{\frac{k-1}{k}} \left[\frac{1}{C(w,w)} - \frac{1}{2C(w,1)} \right]^{\frac{1}{k}}} > \frac{g}{f}.$$

The left hand side is bounded away from zero and infinity as long as $T \leq \frac{C(w,1)}{C(w,w)}$. This range includes the moderate transport cost region. It follows that for g/f small enough the derivative is negative.

References

- [1] Axtell, Robert L. (2001), "Zipf Distribution of U.S. Firm Sizes," *Science*, 293, 1818-1820.
- [2] Brainard, Lael S. (1997) "An Empirical Assessment of the Proximity-Concentration Trade-off between Multinational Sales and Trade," *American Economic Review*, 87, 520-544.
- [3] Ekholm, Karolina, Forslid, Rikard and Markusen, James (2003) "Export Platform Foreign Direct Investment" NBER Working Paper No. 9517.
- [4] Feinberg, Susan E. and Keane, Michael P. (2003) "Accounting for the Growth of MNC-Based Trade Using a Structural Model of U.S. MNCs" University of Maryland, manuscript.
- [5] Grossman Gene M. and Helpman, Elhanan (2002), "Outsourcing in the Global Economy," NBER Working Paper No. 8728, forthcoming in the *Review of Economic Studies*.
- [6] Grossman, Gene M, and Helpman, Elhanan (2003), "Outsourcing versus FDI in Industry Equilibrium," *Journal of the European Economic Association*, 1, 317-327.
- [7] Grossman Gene M. and Helpman, Elhanan (2004), "Managerial Incentives and the International Organization of Production," forthcoming in the *Journal of International Economics*.
- [8] Hanson, Gordon H., Mataloni, Raymond J. and Slaughter, Matthew J. (2001), "Expansion Strategies of U.S. Multinational Corporations," *Brookings Trade Forum 2001*, 245-294.
- [9] Helpman, Elhanan (1984) "A Simple Theory of International Trade with Multinational Corporations," *Journal of Political Economy*, 92, 451-471.
- [10] Helpman, Elhanan and Krugman, Paul (1985) *Market Structure and Foreign Trade*, Cambridge, MA, The MIT Press.
- [11] Helpman, Elhanan, Melitz, Marc and Yeaple, Stephen (2003). "Export versus FDI with Heterogenous Firms," forthcoming in the *American Economic Review*.
- [12] Markusen, James R. (1984) "Multinationals, Multi-Plant Economies, and the Gains from Trade," *Journal of International Economics*, 16, 205-226.
- [13] Markusen, James R. (2002) *Multinational Firms and the Theory of International Trade*, Cambridge, MA and London, UK: The MIT Press.
- [14] Markusen, James R. and Venables, Anthony J. (1998) "Multinational Firms and the New Trade Theory," *Journal of International Economics*, 46, 183-203.

- [15] Markusen, James R. and Venables, Anthony J. (2000) “The Theory of Endowment, Intra-Industry, and Multinational Trade,” *Journal of International Economics*, 52, 209–234.
- [16] Melitz, Marc (2002) “The Impact of Trade on Intra-Industry Reallocations on Aggregate Industry Productivity,” NBER Working Paper No. 8881, forthcoming in *Econometrica*.
- [17] UNCTAD (1998) *World Investment Report: Trends and Determinants*, New York and Geneva: United Nations Conference on Trade and Development.
- [18] UNCTAD (2002) *World Investment Report: Transnational Corporations and Export Competitiveness*, New York and Geneva: United Nations Conference on Trade and Development.
- [19] Yeaple, Stephen R. (2003), “The Complex Integration Strategies of Multinationals and Cross Country Dependencies in the Structure of Foreign Direct Investment,” *Journal of International Economics*, 60, 293-314.