NBER WORKING PAPER SERIES

A SIMPLE THEORY OF MULTINATIONAL CORPORATIONS AND TRADE WITH A TRADE-OFF BETWEEN PROXIMITY AND CONCENTRATION

S. Lael Brainard

Working Paper No. 4269

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 1993

I am grateful to Gene Grossman, Paul Krugman, and Julio Rotemberg for helpful discussions, and to the National Science Foundation for support under grant SES9111649. This paper is part of NBER's research program in International Trade and Investment. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #4269
February 1993

A SIMPLE THEORY OF MULTINATIONAL CORPORATIONS AND TRADE WITH A TRADE-OFF BETWEEN PROXIMITY AND CONCENTRATION

ABSTRACT

This paper develops a two-sector, two-country model, where firms in a differentiated products sector choose between exporting and multinational expansion as alternative modes of foreign market penetration, based on a trade-off between proximity and concentration advantages. The differentiated sector is characterized by multi-stage production, with increasing returns at the corporate level associated with some activity such as R&D, scale economies at the plant level, and a variable transport cost that rises with distance. A pure multinational equilibrium, where two-way horizontal expansion across borders completely supplants two-way trade in differentiated products, is possible even in the absence of factor proportion differences. It is more likely the greater are transport costs relative to fixed plant costs, and the greater are increasing returns at the corporate level relative to the plant level. The model also establishes conditions for a mixed equilibrium, in which national and multinational firms coexist.

S. Lael Brainard
Massachusetts Institute of Technology
Sloan School
Cambridge, Massachusetts 02129
and NBER

I. Introduction

Multinational enterprises (MNEs) are a substantial and imperfectly understood part of the landscape of international trade. Alan Rugman (1988) estimates that the largest 500 multinationals control over one-half of global trade flows, and one-fifth of global GDP. For countries such as the US, which have a well-developed network of overseas affiliates, affiliate sales tend to swamp export flows. DeAnne Julius (1990) estimates that in markets such as Brazil and the EC, local sales by U.S.-owned affiliates are five times the level of imports from the U.S.

It is also striking that a large and growing share of foreign direct investment involves industrialized nations as both the source and destination markets, rather than flowing from North to South. Between 1961 and 1988, over half of all direct investment outflows generated by G-5 countries were absorbed by other G-5 countries; this share had risen to nearly 70 percent by 1988 (Julius (1990)). The share of direct investment absorbed by developed market nations more broadly was 81 percent by the late 1980s (Krugman and Graham (1992)). These trends suggest a large and growing level of two-way activity across borders motivated by considerations of market access. But multinational expansion motivated by market access is not just a recent phenomenon. Indeed, in her study of American multinationals from the 1700s to 1914, Mira Wilkins concludes that market access was the primary motivation for multinational expansion:

> These early international businesses established their foreign stakes - as we have seen - first and foremost in the areas around the world where there were customers - in Europe and Canada primarily. They made their principal investments where there was demand for their offerings - in the wealthiest nations in the world. The businesses described ... went abroad to sell and to make profits through sales in the countries in which they made investments; if they went into manufacturing, refining, or real estate, it was with marketing considerations paramount. They did not go abroad for supplies or to produce abroad for sale in the United States. (Wilkins, 1970, p. 1966)

The implications of these trends for trade and for policy have not been fully explored, partly because research on trade and on multinationals proceeds along separate tracks. Trade theory has made great strides in recent years in explaining the phenomenon of intraindustry trade, and in modelling imperfect competition[1]. Meanwhile, the managerial literature on multinationals has evolved to incorporate recent advances in the theory of the boundaries of the firm.[2] However, there have been only a few attempts at synthesis, even though the characteristics that have been adduced to explain intraindustry trade - imperfect competition, scale economies, and differentiated products - are precisely the characteristics that are associated with the internalization and ownership advantages believed to motivate the formation of multinationals.

This paper develops a two-sector, two-country model, where firms in a differentiated products sector choose between exporting and cross-border expansion as alternative modes of foreign market penetration. The differentiated sector is characterized by increasing returns at the firm level due to some input, such as R&D, that can be spread among any number of production facilities with undiminished value, scale economies at the plant level, such that concentrating production lowers unit costs, and a variable transport cost that rises with distance. The transport cost stands in for a variety of advantages from locating close to the customer, which, according to Wilkins, have been a chief motivation for expansion of US multinationals historically: "investments abroad were made to save on transportation costs and warehousing expenses, to obtain superior customer service and to avoid damage in shipping a perishable product (p. 67)." The decision whether to expand abroad via trade or via investment hinges on a trade-off between these proximity advantages and scale advantages from concentrating production in a single location.

In this model, a pure multinational equilibrium, where two-way intraindustry investment completely supplants merchandise trade in the differentiated sector, is possible even in the absence of factor proportion differences, when proximity advantages are strong relative to concentration advantages. Thus, this model provides a rationale for two-way horizontal expansion across borders, which is distinct from the traditional rationale for vertical expansion based on factor endowment differentials. This model formalizes an explanation of two-way investment in automobile assembly facilities between the US and the EC (eg, Volkswagen and GM), for example - in contrast to factor-based theories of multinational activities, which are better suited to explaining US investments in semiconductor assembly and test facilities in Southeast Asia.

With a simple two-stage production process, there is a pure multinational equilibrium in which two-way trade in invisibles completely crowds out trade in goods, even in the presence of symmetric factor proportions. This equilibrium is more likely the greater are transport costs relative to plant-level scale economies, and the greater are increasing returns at the corporate level relative to those at the plant level. Under the opposite circumstances, a symmetric equilibrium with purely national firms obtains, in which there is balanced two-way trade in differentiated products. For intermediate ranges of transport costs relative to plant-level scale economies and of returns at the corporate level relative to the plant level, there is a mixed equilibrium in which multinational and national firms coexist in the differentiated sector. In a symmetric mixed equilibrium, there is two-way trade in both differentiated products and invisibles.

The incorporation of a third stage of production, such as sales, makes possible an equilibrium in which two-way, intrafirm, intraindustry trade in intermediates replaces trade in final goods. This equilibrium holds when concentration advantages dominate in upstream activities, and proximity advantages dominate downstream.

The presence of multinationals has an ambiguous effect on consumer welfare; consumers benefit from a reduction in transport costs, but there is a concomitant reduction in the number of varieties, all else equal. In the absence of multinationals, they consume a larger quantity of differentiated goods produced locally than of imports, so that there is a home-market bias, whereas with two-way multinational production, they consume equal quantities of home and foreign varieties.

When there are both factor proportions and proximity advantages to overseas expansion, the location of each stage of the production process depends on their relative strength. When factor considerations dominate, single-plant multinationals may emerge, while dual-plant multinationals prevail in the presence of sufficiently strong proximity advantages.

I. i. Related Research

The model combines elements from both the imperfect competition trade literature and the strategy literature on multinationals. In recent years, the strategy literature on multinationals has converged on a conceptual framework based on transactions cost analysis of the boundaries of the firm. Roughly, multinational expansion is explained as a firm's optimal response to the conjuncture of three advantages. The first is an ownership advantage in some unique asset that gives the firm market power and is associated with increasing returns across the firm (network capital, proprietary process technology or product designs, or an established reputational capital); the motivation for expansion is to maximize the returns to this asset. Second, there is an internalization advantage, such that the firm is unable to realize the full value of the asset through the market, due to transactions costs or other market failures. This determines the choice of direct investment over licensing. Third, there is an advantage to locating production near consumers or factors across borders, which makes international expansion more profitable than exporting.

While the first two types of advantages are closely linked to literature on the theory of the firm and on transactions costs, the concept of location advantages is loosely defined, and makes no reference to trade theory.[3] This disjuncture between location advantages and trade theory is surprising in light of the important advances made by trade theorists in recent years in incorporating market power and increasing returns to scale into trade models.

The economics literature on multinationals has gone part way in recent years to incorporate these insights. Broadly, two strands of research have considered multinational activity. The first focuses on the choice between licensing and investing across borders. It models internalization as an endogenous response to a market failure that prevents firms from fully exploiting their market power through the market. Ethier (1986) incorporates the internalization decision into a general equilibrium trade model based on specific factor endowments with a differentiated manufacturing sector. The internalization decision of the firm is a response to imperfections in contracting under uncertainty. The location choice is captured by an ad hoc assumption that downstream distribution must be located in the destination consumption market.

Horstmann and Markusen (1987) focus on the internalization decision in a partial equilibrium framework, where production in the destination market may be chosen over licensing in order to maintain a reputation for quality. Ethier and Markusen (1991) similarly focus on the internalization decision in a partial equilibrium framework. The decision between exporting and overseas production via licensing or investment depends in part on a trade-off between variable transport costs and a fixed production cost, but the general equilibrium implications are not explored.

A second strand of research focuses on the choice between exporting and investing across borders, which hinges on location advantages. Krugman (1983), Helpman (1984), Markusen (1984), Helpman and Krugman (1985), and Horstmann and Markusen (1992) incorporate MNEs into general equilibrium trade models, which start from the premise that firms benefit from internalization due to increasing returns at the level of the firm, and focus attention on the location decision. The model I develop below follows this line of research in focusing on the location decision. But while the Helpman and Helpman, Krugman models explain vertical expansion across borders in terms of factor proportions differences, the model presented in Section II formulates the motivation for overseas expansion in terms of a trade-off between proximity and concentration advantages.[4] The model in Section II integrates the central features of the Horstmann, Markusen and Krugman models in a more general model; the Krugman model[5] develops a differentiated products model that focuses on proximity advantages alone, while the Horstmann, Markusen model[6] incorporates a tradeoff between proximity and concentration advantages in the context of Cournot duopoly.

I briefly summarize the main findings of the factor proportion models here, for purposes of comparison. The Helpman and Helpman/Krugman model incorporates multinationals into a factor proportions trade model, where one sector is characterized by product differentiation and multiple-stage production. There are multiplant economies of scale associated with a firm-specific input which has a public goods character, and production of the input is assumed to be relatively more capital-intensive than production of the final good. In this model, multinationals arise only in the presence of sufficiently great factor endowment differences. When factor endowments are sufficiently similar that factor price equalization obtains in the trade equilibrium, there is no incentive for cross-border investment. When factor prices are not equalized under trade, some of the firms in the differentiated sector locate production of the input in the relatively capital-abundant economy and final good production in the relatively labor-abundant economy, and export back to the capital-abundant economy. Thus, this model explains cross-border investment as a response to factor price differentials, and it predicts that multinational activities will only arise in a single direction within an industry, in single-plant firms. It effectively explains one-way direct investment flows between economies with strong factor proportions differences, but has little to say about two-way intraindustry investment flows between economies with similar factor proportions.

This contrasts sharply with the model I develop below, in which multinational activity can take place in the absence of factor proportion differences and in two directions in the same industry, and is undertaken by multiplant firms. To make the findings directly comparable, I develop a general equilibrium model that closely parallels the Helpman and Helpman, Krugman framework, which also has the virtue of making the findings comparable with a monopolistic competition trade model.

Section II describes the structure of preferences and production in a framework with symmetric factor endowments and a two-stage production process, and derives the equilibrium pattern of trade and investment. Section III elaborates the production structure in the differentiated sector to incorporate an additional stage. Section IV elaborates the model to incorporate factor proportions differences as a motivation for overseas expansion, in addition to the proximity/concentration trade-off. Section V concludes.

II. Two-Stage Production with Symmetric Factor Endowments

There are two neighboring countries, A and B, at a distance D apart. Countries are defined by areas between which factors and consumers are immobile. The analysis will initially assume symmetry in factor endowments and in the distribution of consumers across countries, in order to isolate the role of proximity advantages in stimulating multinational activity.

II. i. Consumption

There are two classes of goods: agricultural goods, which are homogeneous, and manufactured goods, which are differentiated. Consumers have identical, homothetic preferences over classes of goods. This assumption permits separation of the consumption decision into two stages: the allocation of expenditure between the two aggregate goods, and the suballocation of manufacturing expenditure among varieties.

It is further assumed that consumers have identical preferences among varieties of the differentiated good, which take the form of a constant-elasticity-of-substitution subutility function of the Dixit-Stiglitz-Spence type. Denoting the differentiated product Q and the share of expenditure allocated to differentiated goods E_Q, the consumer's second tier utility maximization problem can be represented as:

(1) $$Max \quad u_Q(q_{1A},q_{2A},\ldots,q_{1B},q_{2B},\ldots) \quad s.t. \quad \sum_{a=A}^{B}\sum_{k=1}^{n_a} P_{ka}q_{ka} \leq E_Q$$

where q_{ia} is the quantity of variety i originating in market a (=A,B), P_{ia} is the price of variety i produced in market a, and n_a is the number of varieties from market a. The subutility function is defined over all varieties of good q available at the consumer's location:

(2) $$u_Q(q_{1A},q_{2A},\ldots,q_{1B},q_{2B},\ldots) = [\sum_{a=A}^{B}\sum_{k=1}^{n_a} q_{ka}^{\theta}]^{\frac{1}{\theta}} \qquad \theta = \frac{\sigma-1}{\sigma}, \quad \sigma > 1$$

where σ is the elasticity of substitution among varieties. Solving for the first order conditions yields an optimal consumption plan that includes all varieties, where the share of manufacturing expenditure allocated to each variety is inversely related to the relative price of that variety:

(3) $$q_{ka} = \frac{P_{ka}^{-\sigma}}{\sum_a \sum_k P_{ka}^{1-\sigma}} E_Q \qquad a=A,B$$

Differentiating with respect to the price yields an expression for the elasticity of demand for each variety:

(4) $$-\frac{\partial q_{ka}(.)}{\partial P_{ka}}\frac{P_{ka}}{q_{ka}(.)} = \sigma + \frac{P_{ka}^{1-\sigma}}{\sum_a \sum_k P_{ka}^{1-\sigma}}(1-\sigma) \qquad a=A,B$$

Following Helpman, Krugman, I drop the second term in the elasticity expression. This vastly simplifies the analysis without substantially changing the results, for a sufficiently large number of varieties.[7]

Returning to the first-tier utility maximization problem, total utility for a representative consumer is defined:

(5) $$U = U[u_Y(.), u_Q(.)]$$

where utility from consumption of the agricultural good, $u_Y(.)$, is simply assumed to equal consumption of the agricultural good, Y. Assuming a Cobb-Douglas utility function, the consumer's first-tier utility maximization decision can be represented as:

(6) $$Max \quad Q^{\delta}Y^{1-\delta} \qquad s.t. \quad P_Y Y + \sum_{a=A}^{B}\sum_{k=1}^{n_a} P_{ka}q_{ka} = I$$

where P_Y is the price of the agricultural good, and I is income. Solving for the first order conditions yields expenditure shares for each of the aggregate goods that are independent of income:

(7) $$\frac{\sum_{a=A}^{B}\sum_{k=1}^{n_a} P_{ka}q_{ka}}{I} = \delta \qquad \rightarrow E_Q = I\delta$$

(8) $$\frac{YP_Y}{I} = 1-\delta \qquad \rightarrow E_Y = I(1-\delta)$$

II. ii. Production

There are two factors of production: land, G, which is specific to the agricultural sector, and labor, L, which is used in both sectors. The supply of both factors is symmetric: $L_i=\bar{L}$ and $G_i=\bar{G}$. Given the preferences specified above, the full supply of land is used in the agricultural sector.

The price of the agricultural good will be taken as the numeraire. The production technology in the agricultural sector has constant returns to scale, and perfect competition prevails. Thus, the marginal cost of the agricultural good is constant, and its price in country i is just equal to the marginal cost:

(9) $$1 = C^y(w_i) \qquad i=A,B$$

In addition, the agricultural good is transportable at negligible cost between countries.

The manufacturing sector is characterized by product-specific increasing returns. Adopting a Chamberlinian approach with free entry will yield monopolistic competition, as is familiar from increasing returns trade models. The business system in the manufacturing sector is costlessly separable into two activities: corporate activities and production activities. The corporation performs some activities that have a public good character; they benefit all production facilities equally, and can be spread among any number of facilities without diminishing their value to any one. These activities are unique to each variety. Such activities could include R&D in product design or process technology, advertising, or services such as personnel, treasury, and planning; for simplicity, I will refer to them as R&D. Denoting the level of R&D activity as r, total costs per firm located in market i are:

(10) $$C^r(w_i,r) \qquad \frac{\partial C^r(.)}{\partial r} \geq 0 \qquad i=A,B$$

which are simply assumed to be nondecreasing.

Production operations are characterized by increasing returns at the plant level; there is a fixed cost and a variable cost associated with each manufacturing plant. Production costs for a plant located in market i are:

(11) $$C^Q(w_i,r,q_i) = F(w_i)+V(w_i,r,q_i) \qquad \frac{\partial V(.)}{\partial r} < 0$$

To make the analysis more tractable, it is useful to assume that marginal costs are constant:

(12) $$V(w_i,r,q_i) = V(w_i,r)q_i$$

This assumption is not critical; the results require only that the cost function exhibit bounded returns to scale.

Notice that variable costs fall as the level of corporate activity increases, and that this relationship is independent of the number of plants in operation. The firm chooses r to minimize costs over both markets such that:

(13) $$-\frac{\partial C^r(w_i,r)}{\partial r} = \frac{\partial V(w_i,r)}{\partial r}q_i+\frac{\partial V(w_j,r)}{\partial r}q_j \qquad j=B \text{ when } i=A, \text{ and the reverse}$$

Finally, there is a transport cost. It is intended to stand in for a variety of disadvantages of distance from destination markets, such as shipping costs, trade barriers, linguistic or cultural differences, and slow responsiveness to consumers. The transport cost is assumed to increase with distance: it is modelled as a fraction of output that is lost in transit (similar to Krugman (1979)), which I specify as exponentially increasing in the distance travelled. Thus, for some amount q_a produced in market a=A,B, the amount that survives shipment to a foreign market is smaller by an amount that increases in the distance between the two markets, D, and the transport cost, T:

(14) $$q_a\, e^{-Td_{aj}} \qquad d_{aa}=0,\ d_{ab}=D$$
$$a=A,B,\ b=B \text{ when } a=A, \text{ and the reverse}$$

The competition among firms in the differentiated sector takes place in two stages, and there is a Nash equilibrium in each stage. First, firms in both markets simultaneously decide whether to enter the industry and choose their plant configurations. Then, with full knowledge of competitors' configuration decisions, firms simultaneously choose prices, yielding a Bertrand equilibrium. Solving the first order conditions for profit maximization for a firm located at a(=A,B) yields a price to location j as a markup over the marginal cost of production and transport:

(15) $$P_{aj} = \left[1-\frac{1}{\sigma}\right]^{-1} V(.)e^{Td_{aj}} \qquad \textit{for } a=A,B \textit{ and } j=A,B$$

The markup hinges on σ, the elasticity of substitution between varieties.

II. iii. Market Equilibrium

With constant returns in the agricultural sector, the equilibrium wage is the value of the marginal product of labor:

(16) $$w_i = \frac{\partial Y(\overline{G},L)}{\partial L} \qquad i=A,B$$

With symmetric factor supplies, and identical preferences, the wage is equal across countries.

Together, the markup pricing policy in equation (15) and the elasticity expression in equation (4) imply that all varieties of the differentiated product produced at A will be priced equally at a given location, and similarly for all varieties produced at location B. In market B, the price of varieties produced at A exceeds that of varieties produced at B in market B by an amount that is proportional to the transport cost differential, and the reverse. Given equilibrium prices and numbers of varieties produced in each market, the quantity demanded by a consumer in market a(=A,B) of a representative variety produced at a is:

$$(17)\; q_{aa}(P,n) = \frac{\delta I}{V(.)}\frac{\sigma-1}{\sigma}\frac{1}{\left(n_a+n_b e^{TD(1-\sigma)}\right)} \qquad P=P_{A1},...P_{n_A};P_{B1},...P_{n_B},\; n=n_A,n_B$$

which is increasing in the distance between the markets, and decreasing in the number of varieties available from both markets. The quantity demanded of imported varieties is:

$$(18)\; q_{ba}(P,n) = \frac{\delta I}{V(.)}\frac{\sigma-1}{\sigma}\frac{e^{-TD\sigma}}{\left(n_a+n_b e^{tD(1-\sigma)}\right)} \qquad P=P_{A1},...P_{n_A};P_{B1},...P_{n_B},\; n=n_A,n_B$$

a=B when b=A, and the reverse

The optimal allocation among varieties implies that each consumer consumes equal amounts of all varieties produced at a single location, and smaller amounts of imported varieties, yielding a home market bias.

Depending on the parameter values, there are three possible types of equilibrium: all firms operate two production facilities, all firms operate a single production facility, or there is a mixture of both configurations. I will characterize the conditions for each type of equilibrium in turn. Starting with the equilibrium with single-plant configurations, variable profits on sales of each variety produced at A and sold in A can be expressed as:

$$(19) \qquad \pi^t_{AA}(P_t,n_t) = \frac{\delta I}{\sigma\left(n_A+n_B e^{tD(1-\sigma)}\right)}$$

where t denotes the single-plant equilibrium. Variable profits on sales of each variety produced at A in market B are:

$$(20) \qquad \pi^t_{AB}(P_t,n_t) = \frac{\delta I(e^{tD(1-\sigma)})}{\sigma\left(n_A e^{tD(1-\sigma)}+n_B\right)}$$

Profits are higher the closer is the production location to the destination market, and the fewer competing varieties there are.

Thus, in an equilibrium where all firms have a single production facility, total profits over both markets for a firm located at A are:

$$\Pi_A^t(P_t,n_t) = \pi_{AA}^t(P_t,n_t)+\pi_{AB}^t(P_t,n_t)-F(w)-C^r(w,r) \tag{21}$$

Assuming free entry, the number of firms producing at each location adjusts endogenously to yield zero profits. Further, given the symmetry assumptions made above, the optimal choice of the firm-specific input, r, is symmetric across firms and markets, and the wage in both markets is equal, which in turn implies that the number of firms located at A and B are equal in equilibrium, $n_A=n_B=n_t$. Together, these two conditions imply that the number of firms located in each market is:

$$n_t = \frac{\delta I}{\sigma[F(w)+C^r(w,r)]} \tag{22}$$

The number of varieties is increasing in the expenditure allocated to differentiated products, and decreasing in the elasticity of substitution between varieties, and in the level of per-firm and per-plant fixed costs.

To check whether this is an equilibrium, consider the incentive for a firm to change its production configuration. A firm will open a second production facility only if the increase in its fixed costs is more than compensated by the increase in variable profits from locating closer to foreign consumers. Suppose a firm with a plant located at A opens an additional production facility at B. Prices and profits in A are unchanged. The firm lowers its price in market B, to reflect the reduction in the variable costs of serving that market:

$$P_B^I = \frac{\sigma}{\sigma-1} V(.) \tag{23}$$

yielding variable profits in market B of:

$$\pi_{AB}^I(P_t, n_t) = \frac{\delta I}{\sigma[(n_t-1)e^{tD(1-\sigma)}+(n_t+1)]} \tag{24}$$

Thus, firms are discouraged from defecting only if the increase in variable profits is insufficient to cover the increase in fixed costs:[8]

$$\Delta\pi = \frac{\delta I}{\sigma} \frac{1}{\left[(n_t-1)e^{TD(1-\sigma)}+(n_t+1)\right]} - \frac{e^{TD(1-\sigma)}}{n\left[1+e^{TD(1-\sigma)}\right]} \leq F(w) \tag{25}$$

Putting the no-defection condition together with the zero-profit condition yields the condition for equilibrium:

$$\frac{F(w)}{F(w)+C^r(w,r)} > \frac{(1-e^{TD(1-\sigma)})}{(1+e^{TD(1-\sigma)})} \frac{n_t(1+e^{TD(1-\sigma)})-e^{TD(1-\sigma)}}{n_t(1+e^{TD(1-\sigma)})+1-e^{TD(1-\sigma)}} \tag{26}$$

Comparative statics on the equilibrium conditions establish that the equilibrium in which each firm has a single plant is more likely to prevail the higher is the fixed production cost, the lower are the transport cost and the distance between the markets, and the lower is the corporate fixed cost. In the limit, as the corporate cost goes to 0, a single-plant equilibrium always prevails. The effect of the elasticity of substitution is ambiguous. However, by simplifying the no-defection condition, so that firms do not take into account the effect of their defection on the market price index, it can be shown that the equilibrium is more likely to hold the smaller is the elasticity of substitution. In addition, for a fixed number of firms, a single-plant equilibrium is more likely the smaller is the foreign market.

Next consider an equilibrium in which all firms have production facilities in both markets, denoted m. In this case, a firm serves consumers in market A from its production facility at A, and consumers in B from its production facility at B. Solving for equilibrium prices and quantities yields sales for each variety in each market of:

(27) $$q_{aa}^{m}(P_m,n_m) = \frac{\delta I}{(n_A+n_B)}\frac{\sigma-1}{\sigma} \qquad a=A,B$$

associated with variable per-firm profits in each market of:

(28) $$\pi^{m}(P_m,n_m) = \frac{\delta I}{\sigma(n_A+n_B)}$$

Total per-firm profits over both markets is:

(29) $$\Pi^{m}(P_m,n_m) = 2\pi^{m}(P_m,n_m)-2F(w)-C^{r}(w,r)$$

Again, given symmetry in factor proportions and demand, the number of firms in each market is equal, $n_A=n_B=n_m$. Then assuming free entry, and solving for the number of firms under zero-profits in a dual-plant equilibrium yields:

(30) $$n_m = \frac{\delta I}{\sigma[2F(w)+C^{r}(w,r)]}$$

Again the number of firms is decreasing in the level of total fixed costs, but here the per-plant fixed costs are weighted more heavily. Notice that under plausible conditions, there are fewer varieties in total in the dual-plant equilibrium, regardless of the level of transport costs.[9]

A firm is tempted to defect from the dual-plant equilibrium if the savings in fixed costs from shutting down a production facility more than offset the loss in variable profits due to higher transport costs to the foreign market. Consider a firm that shuts down its plant at B. Pricing and profits in A are unchanged. In B, however, the firm's price rises to reflect the increase in transport costs:

(31) $$P' = \frac{\sigma}{\sigma - 1} V(.) e^{TD}$$

yielding lower variable profits of:

(32) $$\pi'_{AB}(P', P_m, n_m) = \frac{\delta I}{\sigma} \frac{e^{TD(1-\sigma)}}{[(2n_m - 1) + e^{TD(1-\sigma)}]}$$

Thus, firms have no incentive to defect from a dual-plant configuration only if:

(33) $$F(w) \leq \frac{\delta I}{\sigma}\left(\frac{1}{2n_m} - \frac{e^{TD(1-\sigma)}}{(n_m - 1) + e^{TD(1-\sigma)}}\right)$$

Combining the zero-profit and no-defection conditions yields equilibrium conditions:

(34) $$\frac{2F(w)}{2F(w) + C'(w,r)} < \frac{2n_m(1 - e^{TD(1-\sigma)}) - (1 - e^{TD(1-\sigma)})}{2n_m - (1 - e^{TD(1-\sigma)})}$$

A fully multinational equilibrium, where all firms have production facilities in both markets, is more likely the smaller is the fixed production cost, the higher are transport costs and the distance between markets, and the higher are economies at the corporate. Here, as the corporate fixed cost goes to 0, a dual-location equilibrium is never sustainable. Again, the effect of an increase in the elasticity of substitution between varieties is ambiguous. By simplifying the no-defection condition as above, it can be shown that an increase in the elasticity of substitution makes the multinational equilibrium more likely to hold. For a fixed number of firms, the dual-plant equilibrium is more likely to hold the larger is the foreign market. Thus, the dual-plant and single-plant equilibria are characterized by precisely reversed conditions.

In the intermediate range of parameter values, there is a third possible equilibrium, in which multinational firms coexist with national firms. In the mixed equilibrium, some fraction, α, of firms have a single production facility and export, and the remaining fraction have production facilities in both markets.

The condition for the mixed equilibrium can be derived from combining no-defection conditions for both single plant and dual-plant firms:

$$n_c(2-\alpha(1-e^{TD(1-\sigma)})) > 1+e^{TD(1-\sigma)} \qquad (35)$$

where n_c is the number of firms in the mixed equilibrium. When returns to the two configurations are equalized, the expression for the proportion of single plant firms is:

$$\alpha = \frac{2}{(1-e^{TD(1-\sigma)})} - \frac{\delta I}{\sigma n_c \left(F(w)+C^r(w,r_2)-C^r(w,r_1)\right)}$$

where r_i is the profit-maximizing level of R&D input for a firm with i plants. For a given number of firms, the proportion with production facilities in both markets declines with the size of the fixed production cost and the additional R&D investment, and rises with the size of the transport cost and the distance between markets. With free entry, this equilibrium holds only when the following conditions on the fixed costs and the transport costs are satisfied:

$$F(w) = 2C^r(w,r_1)-(1-e^{TD(1-\sigma)})C^r(w,r_2)$$

The fixed production costs must be greater the greater is the size of the R&D investment for a single-plant firm relative to a dual-plant firm, and the smaller are the transport cost, distance, and the elasticity of substitution. If this condition is satisfied, any mixture of firm types can be supported.

II. iv. Factor Market Clearing

It is now possible to go back and solve for equilibrium conditions in the labor market. Constant returns in the agricultural sector imply that the per unit labor input is simply the derivative of the unit cost function with respect to local wages. Thus, in market i=A,B:

$$l^{y}(w_i) = \frac{\partial C^{y}(w_i)}{\partial w} \tag{36}$$

In the differentiated sector, the demand for labor can be separated into corporate and production activities. The demand for labor for corporate activities is the derivative of the cost function with respect to the local wage at the equilibrium level of corporate services:

$$L^{r}(w_i,r) = \frac{\partial C^{r}(w_i,r)}{\partial w} \tag{37}$$

Production activities generate a demand for labor associated with the fixed production cost, which is simply the derivative with respect to the local wage:

$$L^{F}(w_i) = \frac{\partial F(w_i)}{\partial w} \tag{38}$$

And they generate a demand for labor for production and transport of the differentiated goods; for a unit of output produced at location i and sold in j this demand is:

$$l^{Q}(w_i,r) = e^{Td_{ij}}\frac{\partial V(w_i,r)}{\partial w_i} \qquad d_{ii}=0,\ d_{ik}=D \tag{39}$$

$i=A,B$; $k=B$ when $i=A$, and the reverse

Labor demand varies according to the type of equilibrium in the differentiated sector. With symmetric wages across countries, in the single-plant equilibrium, factor clearing conditions in each market are:

(40) $$\bar{L} = l^{Y}(w)Y^{t}+n_{t}[L^{r}(w,r)+L^{F}(w)+l^{Q}(w,r)q^{t}]$$

where q^t is per firm output in the single-plant equilibrium. In the pure dual-plant equilibrium, factor clearing implies:

(41) $$\bar{L} = l^{Y}(w)Y^{m}+n_{m}[L^{r}(w,r)+L^{F}(w)+2l^{Q}(w,r)q^{m}]$$

where q^m is per firm output in the dual-plant equilibrium. In the mixed equilibrium, factor demands are straightforward to calculate, and depend on the mix of firms.

II. v. Commodity Market Clearing

The income-expenditure equality can then be expressed in market A in the dual-plant equilibrium as:

(42) $$I^{m} = Y+2n_{m}[P_{A}^{m}q_{A}^{m}(P^{m})]$$

while in the single-plant equilibrium it is:

(43) $$I^{t} = Y+n_{t}[P_{A}^{t}q_{A}^{t}(P^{t})+P_{B}^{t}q_{B}^{t}(P^{t})]$$

Then commodity market clearing requires a single market clearing condition. For agriculture in equilibrium n this is:

(44) $$(1-\delta)I^{n} = Y \qquad n=m,t,c$$

II. vi. Trade Balance

Given symmetry in population size and income, there are symmetric demands in both markets for the agricultural good and for differentiated products. The single-plant equilibrium is essentially a trade equilibrium with purely national firms. Given symmetric factor proportions, each country satisfies its own demand for the agricultural good, and there are balanced flows of two-way trade in the differentiated manufacturing sector:

$$\frac{\delta I e^{-TD\sigma}}{\sigma(1+e^{TD(1-\sigma)})}$$

The volume of intraindustry trade is a decreasing function of the transport cost, the distance between markets, and the elasticity of substitution, and an increasing function of income and the expenditure share allocated to differentiated goods.

In sharp contrast, there is no trade in goods in the pure multinational equilibrium. Instead of two-way trade in goods in the differentiated sector, there is balanced two-way trade in corporate services. The flow in each direction is:

$$n_m w L^{r}(w,r)$$

It rises with the amount of labor employed in corporate activities and with the number of firms, which in turn is an increasing function of the expenditure allocated to differentiated products, and a decreasing function of the elasticity of substitution and fixed costs.

Thus, this model yields an equilibrium in which multinational production arises in the absence of factor proportion differences, and two-way intraindustry trade in invisibles completely crowds out intraindustry trade in goods. These results contrast sharply with the predictions from a model based on factor proportions differences, where the only motivation for cross-border expansion is factor price differentials caused by different factor endowments. In a factor proportions framework, the presence of multinationals reduces the share of total trade accounted for by intraindustry trade, but need have no effect on the total volume of trade, since varieties headquartered in the capital-intensive country and produced in the labor-intensive country are exported back to the capital-intensive country.

Lastly, in the mixed equilibrium, there is balanced two-way trade in both goods and invisibles in the differentiated sector. Multinational sales are a relatively larger share of total foreign sales for each country, the higher the proximity advantage relative to the concentration advantage.

III. Three-Stage Production with Symmetric Factor Endowments

So far, the model suggests the stark result that, given large benefits to proximity relative to benefits from concentration, multinational enterprises will either in part or totally substitute for intraindustry trade. This sits uneasily alongside two stylized facts. First, there is a large share of intraindustry trade in total trade (Loertaches, Walter (1981)), and second, multinationals account for a significant share of trade (Rugman (1988)). Indeed, there is evidence that suggests that multinational activity on balance is complementary to rather than a substitute for trade (Swedenborg (1979), Blomstrom et. al. (1984)).

The first fact could be explained in the model developed above by a combination of low benefits to proximity and large scale economies. The second fact is not explained by the model developed above with a simple two-stage production function in the differentiated sector. A factor proportions account of multinationals would explain the second fact in two ways. First, given a two-stage production process and single-plant multinationals, finished goods are exported back from the country of production to the headquarter country. Second, with a three-stage production process, multinationals produce intermediate inputs in the economy that is relatively abundant in the factor used relatively intensively in intermediate production, and export them to downstream production facilities in the economy that is relatively well-endowed with the factor used relatively intensively in final goods processing. A factor proportions model thus predicts that either final goods or intermediates will flow across borders in a single direction, contributing to interindustry rather than intraindustry trade.

Alternatively, the model developed above can be elaborated in a way that is sensible and accords well with both stylized facts. By elaborating the production structure in the differentiated products industry to incorporate additional stages characterized by different proximity/concentration trade-offs, the model predicts that multinationals will generate intraindustry, intrafirm trade under plausible conditions. This prediction differs from a factor proportions account, which would predict one-way flows of intrafirm trade.[10]

Again, the contrast between commodity semiconductors and cars serves to illustrate the differences in the two models. In the 1970s, US commodity chip firms made substantial investments in assembly and test facilities in Southeast Asia, which were motivated entirely by factor price differentials.[11] The test and assembly stages were intensive in low-skilled labor, in contrast to the design and manufacturing stages, which were intensive in engineering skills and capital. Semiconductors were shipped from the US design and manufacturing complexes to assembly and test facilities in Southeast Asia, and then shipped back to sales destinations in the US and elsewhere.[12]

In contrast, in the automobile industry, there has been substantial two-way investment and multinational expansion into destination markets. A typical pattern of expansion would start with the establishment of distribution and sales facilities in the destination market, followed by assembly facilities when the share in the foreign market was sufficient to warrant the fixed investment.[13] Factor price differentials provide little insight into these investment patterns, since the markets in question, such as the EC and the US, have had similar relative factor prices in periods of substantial two-way multinational activity. Instead, the relationship between scale economies and proximity advantages provides more insight: scale economies are substantial but bounded in auto assembly, and smaller in sales and distribution, trade barriers into several countries in the EC are high, and the advantages of a local presence and customization to the local market are high.

To formalize an explanation for the latter pattern, I modify the production structure to incorporate an additional stage, which has a distinct proximity/concentration trade-off.[14] Thus, in addition to the R&D (r) and manufacturing (q) stages detailed above, there are sales activities (s), whose production requires a fixed cost and a constant marginal cost:

$$C^{s}(w_a,q,s) = F^{s}(w_a)+V^{s}(w_a,q)s \qquad a=A,B \tag{45}$$

Similar to the upstream relationship between corporate activities and manufacturing, the variable cost in sales is a decreasing function of the input from the manufacturing stage, which is proprietary to the firm. Unlike the R&D input, however, the manufactured input lowers the cost only of the sales facility that uses it.

In addition, there is a transport cost, both for the manufactured input, and for the final output. To simplify notation, I replace the exponential transport cost with a more general parameter, t, which subsumes both the distance between markets and the shipping cost. The cost of transporting the output of each stage, $t_i > 1$ ($i=q,s$), is the reciprocal of the fraction lost in shipment; only the portion of output that is shipped across borders incurs the transport cost. The transport cost between manufacturing and sales can be interpreted in terms of a co-location advantage.

Each firm chooses its production configuration to trade off additional fixed costs against additional variable costs of transport at each stage of production. Casual empiricism suggests that sales activities are most commonly dispersed near consumers, while manufacturing is more likely to be concentrated, and corporate activities such as R&D are the most concentrated. Accordingly, as in Section II, I assume that R&D output can be transported costlessly to manufacturing plants ($t_r = 1$).

Again, firms compete in two stages: Nash equilibrium in production configurations is followed by Nash equilibrium in prices. Given symmetry in factor proportions and demand, factor prices are equal in equilibrium, $w_A = w_B$, and the number of firms (or R&D investments) is equal in markets A and B, so we can confine attention to firms headquartered in market A without loss of generality.

There are four possible configurations in the differentiated sector. Each firm can concentrate both manufacturing and sales in a single market; it can invest in sales facilities in both markets, and concentrate manufacturing in a single market; it can spread both sales and manufacturing activities across borders; or it can invest in manufacturing facilities in both markets, and concentrate sales in a single market. The fourth configuration can be ruled out, since the firm would save both fixed production costs and transport costs by shutting down a manufacturing facility. In any configuration, R&D activities are concentrated in a single location, due to the combination of scale economies and costless transport.[15]

I start by deriving conditions for each pure type of equilibrium - where all firms headquartered in each market have identical configurations.[16] The equilibrium conditions are derived in precisely the same manner as in Section II, by combining zero-profit conditions with no-defection conditions for all firms located in each market. However, with three possible configurations, there are two no-defection conditions for each equilibrium.

Firms choose configurations and prices taking as given the prices and configurations of their rivals. Variable profits for each firm across both markets in equilibrium j are:

(46) $$\pi^j(P_A^j,P_B^j,n_j) = \frac{\delta I}{\sigma n_j} \qquad j=t,m,d$$

where n_j is the number of firms located in each market under market structure j.

The number of firms in each equilibrium is determined endogenously by the zero-profit condition under free entry. Thus, in an equilibrium where firms concentrate all activities in a single location (denoted t), it is:

$$(47) \qquad n_t = \frac{\delta I}{\sigma}[C^r(w,r)+F^q(w)+F^s(w)]^{-1}$$

Just as in Section II, in each equilibrium the no-defection conditions are obtained by comparing the change in variable profits associated with each alternative production configuration with the corresponding change in fixed costs. The equilibrium is then obtained by solving the no-defection and zero-profit equations simultaneously, yielding conditions on the relative size of the transport costs to the fixed costs, and on the activity-specific scale economies relative to the firm-wide scale economies.[17] Combining the condition ruling out defection to a fully diversified configuration with the zero-profit condition yields an equilibrium condition:

$$(48) \qquad \frac{F^s(w)+F^q(w)}{C^r(w,r)} > \frac{[n_t(1+t_s^{1-\sigma})-t_s^{1-\sigma}]}{(1+2t_s^{1-\sigma})}$$

where the cost of transporting intermediate inputs raises variable costs by a proportion, T_q, defined:

$$T_q = \frac{(V_q(.)t_q+V_s(.))}{(V_q(.)+V_s(.))}$$

And the equilibrium condition ruling out defection to a downstream diversified configuration is:

$$(49) \qquad \frac{F^s(w)}{C^r(w,r)+F^q(w)} > \frac{(T_q^{1-\sigma}-t_s^{1-\sigma})[n_t(1+t_s^{1-\sigma})+t_s^{1-\sigma}]}{n_t(1+t_s^{1-\sigma})(1+2t_s^{1-\sigma}-T_q^{1-\sigma})+(T_q^{1-\sigma}-t_s^{1-\sigma})(1+2t_s^{1-\sigma})}$$

In the equilibrium where only downstream activities are dispersed across borders (denoted d), the equilibrium number of firms is:

(50) $$n_d = \frac{\delta I}{\sigma}[C^r(w,r)+F^q(w)+2F^s(w)]^{-1}$$

There are two no-defection conditions, corresponding to the two alternative configurations. Combining the zero-profit condition with the condition ruling out defection to a single-location configuration yields:

(51) $$\frac{(T_q^{1-\sigma}-t_s^{1-\sigma})[n_d(1+T_q^{1-\sigma})-T_q^{1-\sigma}]}{n_d(1+T_q^{1-\sigma})(1+2t_s^{1-\sigma}-T_q^{1-\sigma})-(T_q^{1-\sigma}-t_s^{1-\sigma})(1-T_q^{1-\sigma})} > \frac{F^s(w)}{C^r(w,r)+F^q(w)}$$

And combining the zero-profit condition with the condition that firms have no incentive to extend their multinational activities by opening an additional manufacturing facility yields:

(52) $$\frac{F^q(w)}{C^r(w,r)+2F^s(w)} > \frac{(1-T_q^{1-\sigma})[n_d(1+T_q^{1-\sigma})-T_q^{1-\sigma}]}{n_d(1+T_q^{1-\sigma})2T_q^{1-\sigma}+(1-T_q^{1-\sigma})(1+2T_q^{1-\sigma})}$$

Lastly, the number of firms in the equilibrium where firms diversify both manufacturing and sales across borders (denoted m) is:

(53) $$n_m = \frac{\delta I}{\sigma}[C^r(w,r)+2F^q(w)+2F^s(w)]^{-1}$$

The combination of the zero-profit condition with the condition that firms have no incentive to close down one production facility yields the equilibrium condition:

(54) $$\frac{(2n_m-1)(1-T_q^{1-\sigma})}{4n_m T_q^{1-\sigma}} > \frac{F^q(w)}{C^r(w,r)+2F^s(w)}$$

And the equilibrium condition ruling out defection to a concentrated, single-location configuration is:

$$(55) \qquad \frac{(2n_m-1)(1-t_s^{1-\sigma})}{2n_m t_s^{1-\sigma}} > \frac{F^s(w)+F^q(w)}{C^r(w,r)}$$

Notice that together, equations (47), (50), and (53) imply that under plausible conditions, the number of varieties is greatest in the single-location equilibrium, and smallest in the fully diversified dual-location equilibrium.[18]

Comparative statics on the equilibrium conditions establish that a fully diversified equilibrium, where firms establish sales and manufacturing facilities in both markets, is more likely the smaller are fixed costs in manufacturing, the greater are fixed costs in R&D, the greater are the costs of transporting both the intermediate input (T_q)[19] and the final good (t_s), and the greater is the elasticity of substitution.[20] A fully concentrated, single-location equilibrium is more likely the greater is the fixed cost in sales, the smaller is the fixed cost of R&D, the lower is the transport cost of the final good relative to that of the intermediate input, and the smaller is the elasticity of substitution. A downstream-diversified equilibrium is more likely to obtain the smaller is the fixed cost in sales, the greater is the fixed cost in manufacturing, and the smaller is the cost of transporting the intermediate input relative to that of the final good. In addition, the downstream-diversified equilibrium is only possible when the transport cost of the intermediate is smaller than that of the final good. Similar to the two-stage case, a diversified equilibrium is increasingly likely the larger are the returns to scale at the corporate level relative to those in manufacturing and sales.

The pattern of trade and production associated with each equilibrium can be derived following the analysis in Section II, so that I will simply discuss the main conclusions here, without detailing the factor and commodity clearing conditions. The addition of a stage of production generates demand for labor, analogous to equations (38) and (39). Conditions for factor market clearing and commodity market clearing can then be derived analogously to equations (40) through (44), with a similar set of conditions for the downstream diversified equilibrium. Just as in Section II, with equal relative factor endowments, identical production patterns will emerge in the two markets; given symmetric demand, trade in the differentiated sector will be balanced, and the agricultural sector will be autarkic.

The equilibrium in which all firms concentrate all three production stages in a single market results in trade patterns identical to those in a pure trade model with differentiated products and symmetric factor proportions: there are balanced flows of two-way trade in differentiated products, and no interindustry or intrafirm trade. Just as in a trade model, the flows will be greater, the larger is the share of expenditure allocated to differentiated products, the higher is income, and the lower are transport costs.

The equilibrium in which all firms invest in both manufacturing and sales facilities in both markets results in balanced two-way flows of intrafirm trade in R&D completely crowding out trade in goods. Here multinational production completely substitutes for trade.

The addition of a production stage introduces the possibility of a third equilibrium, in which firms diversify their sales activities across borders but concentrate their production activity in a single market. In this equilibrium, there are balanced two-way flows of intraindustry, intrafirm trade in intermediates, which substitute completely for trade in final goods and in invisibles. This equilibrium is thus consistent with both of the stylized facts cited above: the existence of multinationals generates intraindustry trade.

IV. Asymmetric Factor Proportions

In reality, firm configurations are likely to reflect considerations of both factor and proximity advantages; in this section I develop a hybrid model to accommodate both considerations. As described above, a pure factor proportions account of MNEs predicts configurations in which a firm locates R&D in the country relatively abundant in the factor in which R&D is relatively intensive, and a single manufacturing plant in the country abundant in the factor in which manufacturing is intensive. Such accounts use as their benchmark a world in which factors are completely mobile - the integrated equilibrium - and analyze the implications for firm configurations and trade patterns of dividing up the supply of factors. It is somewhat complicated to use this approach in analyzing a hybrid model of multinationals, since there is no obvious counterpart to the integrated equilibrium when transport costs are introduced. Instead, I adopt the symmetric equilibrium as an alternative benchmark and analyze how firm configurations and trade patterns change as factor proportions become increasingly unbalanced.

This section returns to the simplified two-stage production structure of the first section. Two factors are assumed to be used in the production of both goods: skilled labor (L_1) and unskilled labor (L_2). The factors are divided between the two countries and immobile. Type j is paid wage w_j^i in country i. Define the ratio of wages of skilled workers to that of unskilled workers as $W^i = w_1^i/w_2^i$. Next, define the equilibrium unit input for skilled labor in activity k as $l_1^k(W)$, given relative wages W, and similarly for unskilled labor. Further, assume that the factor intensity of production in the three productive activities is ranked consistently at any relative wage rate, with R&D the most skill-intensive activity and agriculture the least: $l_1^r(W)/l_2^r(W) > l_1^q(W)/l_2^q(W) > l_1^y(W)/l_2^y(W)$.

In addition, assume that there is some small advantage to co-location, such that a firm chooses to co-locate its first manufacturing plant with its R&D facilities, all else equal.[21] In the equilibrium with equal factor proportions, the existence of a co-location advantage, however small, eliminates single-plant multinationals.

Starting in the benchmark case with equal factor endowments, there are three possible equilibria: a fully multinational equilibrium where all firms in the differentiated sector have manufacturing facilities in both markets, a pure trade equilibrium where all firms have a single manufacturing plant, and a mixed equilibrium. In all three cases, given equal factor proportions, factor prices are equalized across the two markets, and there is symmetry in the number of firms. Suppose that the aggregate ratio of skilled to unskilled labor is $\bar{L}_1/\bar{L}_2$. We will analyze the effect of moving skilled labor from market B to market A and unskilled labor from market A to B in the same proportions as the aggregate skilled to unskilled ratio, so that in the limit economy A is the monopoly supplier of skilled labor, and B is the monopoly supplier of unskilled labor. This corresponds to movement from the center along the diagonal toward the northwest corner in the Helpman diagram.

In the pure trade equilibrium, as the ratio of skilled to unskilled labor rises in A and falls in B, there is downward pressure on the relative wage of skilled labor in A, and upward pressure in B. Firms respond to the imbalance by moving R&D activities to A, while agriculture expands in B, thereby maintaining factor price equalization. In this case, firms adopt multinational configurations in response to factor availabilities; these configurations are distinct from those found in the symmetric multinational equilibrium in that each firm has a single plant. Here there are two-way flows of differentiated products of the same overall magnitude as in the symmetric equilibrium, but these flows may be unbalanced, reflecting income differentials, and there are additional trade flows of agriculture from B to A in return for R&D services from A to B.

As reallocation of factor supplies continues, so that the relative abundance of skills in A intensifies, eventually one country becomes the sole producer of agriculture or R&D. At that juncture, factor prices diverge, and different production techniques are adopted in the two countries. Assume that R&D becomes concentrated before agriculture. Take the point at which the last R&D facility has moved to A, and transfer an additional increment of skilled labor from B to A and of unskilled labor from A to B. Now a manufacturing plant will move to A from

B, and agriculture will continue to move to B from A. With sufficiently strong factor imbalances, either agriculture concentrates completely in B and/or R&D in A, and manufacturing is split between the two in unequal proportions and utilizing different production techniques. Here, there is unbalanced, two-way trade in differentiated products and one-way flows of R&D from A to B and agriculture from B to A.

The interesting feature of this equilibrium is that firms continue to use trade as the sole means of foreign penetration. With unequal factor prices, the payoff to concentrating manufacturing activities must be at least as great as it is with equal factor prices. Thus, the introduction of factor imbalances gives rise only to single-plant multinationals, if any.

Performing the same experiment starting from a pure multinational equilibrium, the initial response to a progressive unbalancing of factor proportions is similar. R&D moves to market A, agriculture moves to market B, and factor price equalization is maintained, along with plant configurations. The pattern of trade adjusts, so that there are flows of agriculture from B to A, balanced by reverse flows of R&D services.

Eventually, the imbalance in factor endowments becomes sufficiently extreme that one industry concentrates in a single market. Now a factor price differential between the two markets is added to the scale economies motivation for concentrating manufacturing activities in a single location. Thus, a few firms will concentrate production in the market with favorable factor conditions, and export, while others will maintain production facilities in both markets. In equilibrium, the returns to both configurations will be equal. So here, although the total extent of multinational activity may remain unchanged, the number of firms with multiplant configurations will diminish, and there will be unbalanced two-way trade in differentiated products along with flows of R&D services and agriculture.

Thus, the introduction of different factor proportions across the two markets permits a broader range of possible trade and production patterns. With unequal factor proportions, the location of each stage of the production process depends on two considerations: the relative concentration/proximity trade-off and relative factor intensities. Unequal factor proportions introduce the possibility of single-plant multinationals, and make the likelihood of single-plant firms higher outside the range of factor price equalization.

V. Conclusion

The introduction of an explicit trade-off between scale economies and proximity advantages in a differentiated products model of trade permits a much richer set of predictions about the pattern of trade and investment. In particular, consideration of this trade-off introduces the possibility of two-way, horizontal multinational activity, even in the absence of factor price differentials, and of intrafirm trade in intangibles completely supplanting trade in goods. It also introduces the possibility of multi-plant configurations across borders. These results, which appear to be of increasing importance empirically, are inexplicable in an account of multinational expansion driven purely by factor proportions differentials.

An equilibrium characterized by multinational production and multiplant configurations is more likely the higher are returns to scale at the corporate level relative to the plant level, the higher are transport costs across markets, the greater is expenditure on differentiated products in the foreign market, and the higher is the elasticity of substitution between varieties.

This framework also can explain equilibria in which national firms coexist with multinational firms in the same industry. A mixed equilibrium is sustainable for intermediate ranges of the parameter values, when returns to the two configurations differ by less than the differential between the associated fixed costs.

Consumption patterns vary with the type of equilibrium in the differentiated sector. In a pure multinational equilibrium, equal quantities of all varieties are consumed, regardless of their origin, while in a pure trade equilibrium, there is a home market bias in the consumption of the differentiated product. Consumers benefit from the presence of multinationals, due to the reduction of the price of imported goods, but there is a concomitant loss in consumer surplus from a reduced overall number of varieties.

The prediction that intraindustry flows of intangibles substitute for trade depends on the structure of production in the differentiated sector and the relative proximity/concentration tradeoff of each stage of production. With multiple production stages, if concentration advantages dominate in upstream production activities, while proximity advantages dominate downstream, then intraindustry, intrafirm flows of intermediates supplant intraindustry flows of final goods. In this case, multinational activity is complementary to trade in goods, and multinationals generate intraindustry trade.

When both factor proportions differences and proximity advantages are incorporated within the same framework, the pattern of trade and investment depends on their relative strength. When concentration advantages are high relative to proximity advantages, the introduction of moderate factor proportions differences results in the formation of single-plant multinationals, and there are unbalanced one-way flows of trade in agriculture and in R&D services in the reverse direction. With more extreme factor proportion differences, factor price differentials arise, and flows of trade in differentiated products become increasingly unbalanced, as manufacturing facilities are distributed increasingly unevenly across the two markets. When proximity advantages dominate, moderate factor proportions differences result in unbalanced one-way flows of trade in agriculture in return for R&D services, but firms maintain multiplant, multinational configurations. As factor proportions become increasingly unbalanced, the resulting factor price differentials create an incentive to concentrate production in a single plant.

This model has strong implications for policy. Most obviously, the imposition of a tariff raises transport costs, and makes penetration by multinational production more attractive, while a tax on the earnings of intangibles (or profits) has the reverse effect. Further, the relative structure of tariffs on final goods and intermediate inputs influences firms' decisions about which activities to concentrate and which to disperse across borders.

The model further implies that restrictions on foreign direct investment may be more damaging to welfare than restrictions on trade in industries where natural proximity advantages are high and scale economies at the corporate level are high relative to those at the plant level, and suggests conditions under which regional trade blocs are not deleterious to global welfare. If regions are geographically coherent areas within which natural barriers are low and between which they are high, and if different regions have similar region-wide factor proportions (while the factor proportions of countries within regions could differ significantly), then the welfare gains from free trade may be exhausted within regions, while free interregional investment flows raise welfare by permitting maximal returns to investment at the corporate level, and a greater variety of products. Future research will analyze these conditions.

Further, this model lends itself to testable hypotheses. In particular, it suggests that firms' observed choices between exports and multinational expansion as alternative modes of foreign market penetration should reflect trade-offs between measurable proximity and concentration advantages. Research currently underway employs a gravity equation approach to test this hypothesis.

1. Seminal contributions to this large and growing literature include Krugman (1979), Lancaster (1980), Helpman (1981), and Ethier (1982).

2. See Caves (1982) and Dunning (1988).

3. This may be attributable in part to the failure of earlier attempts to explain multinational activity in terms of capital flows in a factor proportions trade framework with perfect competition.

4. In the Markusen model, factor endowments are symmetric, but sector-specific capital plays a key role in explaining multinational production.

5. In the model below, firms incur an additional fixed cost to open production facilities abroad, and save on variable transport costs, while in the Krugman model, there is no fixed production cost, and variable production costs differ between markets by some multiplicative factor. The predictions differ in two respects. First, in the Krugman model, only uniform equilibria are possible, while here there is a mixed equilibrium. Second, in the model below, the decision whether to produce abroad depends on the potential scale of production, while in the Krugman model, firms do not discriminate between markets based on size or share.

6. The Horstmann, Markusen model assumes that there are 2 potential entrants, one headquartered in each country, and there is Cournot-Nash equilibrium if both enter the market. Below, the number of firms is determined endogenously in a monopolistic competition framework. This framework lends itself more readily to predictions about the pattern of trade, and introduces an additional welfare effect from the number of varieties produced.

7. The analysis would be messier in the absence of this simplification, but the qualitative results would not change significantly.

8. The no-defection conditions reflect the additional costs that are required to open a production facility, holding fixed the level of R&D input. This can be interpreted either in terms of a sequential decision-making process, or a fixed R&D requirement.

9. Assuming that the R&D investment for a single-plant configuration does not exceed that for a dual-plant configuration by more than the size of the fixed production cost.

10. Unfortunately, the empirical work in this area does not distinguish between the two types of trade when referring to multinationals.

11. See Yoffie (1992) for a complete account of investment and trade patterns in the global semiconductor industry.

12. The current pattern of trade and investment in the semiconductor industry is probably best explained by a hybrid model (see Section IV).

13. In the interests of tractability, I ignore dynamic considerations in this model. Future research will model the relationship between exports and multinational production in a dynamic setting.

14. The model can be generalized to any number of production stages, where each stage i located in country a is assumed to entail a fixed cost, a variable production cost, and a multiplicative transport cost for each unit that is exported:

$$C^i(w_a, q_{i-1}, q_i) = F^i(w_a) + V^i(w_a, q_{i-1}, q_i)\left(\frac{q_i^a + t^i(q_i - q_i^a)}{q_i}\right)$$

where the variable cost is a decreasing function of the output from the previous stage used as an input, q_{i-1}, and internally produced inputs are proprietary.

15. I assume it is always optimal for a firm to locate its first manufacturing plant in the same market as R&D, due to some arbitrarily small co-location advantage between R&D and manufacturing. This assumption is useful because it makes the flows of invisibles determinate, and sensible because it accords well with casual empiricism. It could be modelled formally as $t_r > 1$.

16. I confine attention to uniform equilibria for simplicity; mixed equilibria can be derived similarly to Section II.

17. Similar to Section II, the no-defection condition excludes changes in the R&D investment.

18. Assuming that the R&D investment does not decrease with the number of production and sales locations, or, if it does, that the increase in the fixed costs of production or sales is at least as great as the R&D savings.

19. T_q is more accurately described as the ratio of the variable cost gross of the cost of transporting the input to the variable cost net of transport costs.

20. This result is derived by differentiating the no-defection conditions, simplified so that firms do not take into account the effect of their configuration decisions on the market price index. This is consistent with the assumption of a large number of firms.

21. I simply follow Helpman in making an ad hoc assumption. This can be formalized in a variety of ways, such as assuming a variable transport cost between R&D and manufacturing, or a lower fixed cost for adjacent facilities.

References

Blomstrom, Magnus, Robert Lipsey, and Ksenia Kulchycky (1985): "U.S. and Swedish Direct Investment and Exports," in Robert Baldwin (ed.), Trade Policy and Empirical Issues (Chicago: University of Chicago Press).

Caves, Richard (1982): Multinational Enterprise and Economic Analysis (Cambridge: Cambridge University Press).

Dunning, John (1988): Explaining International Production (London: Harper Collins).

Erdilek, Asim (1985): Multinationals as Mutual Invaders (New York: St. Martin's Press).

Ethier, Wilfred (1982): "National and International Returns to Scale in the Modern Theory of International Trade," American Economic Review, 72, pp. 389-405.

Ethier, Wilfred (1986): "The Multinational Firm," Quarterly Journal of Economics, November, pp. 806-33.

Ethier, Wilfred, and Henrik Horn (1990): "Managerial Control of International Firms and Patterns of Direct Investment," Journal of International Economics, 28, pp. 25-45.

Ethier, Wilfred, and James Markusen (1991): "Multinational Firms, Technology Diffusion, and Trade," NBER Working Paper No. 3825, August.

Helpman, Elhanan (1981): "International Trade in the Presence of Product Differentiation, Economies of Scale, and Monopolistic Competition," Journal of International Economics, 11, pp. 305-40.

Helpman, Elhanan (1984): "A Simple Theory of International Trade with Multinational Corporations," Journal of Political Economy, p. 451.

Helpman, Elhanan, and Paul Krugman (1985): Market Structure and Foreign Trade (Cambridge: M.I.T. Press).

Horstmann, Ignatius, and James Markusen (1987): "Licensing vs. Direct Investment: A Model of Internalisation by the MNE," Canadian Journal of Economics, August, p. 464.

Horstmann, Ignatius, and James Markusen (1992): "Endogenous Market Structures in International Trade," Journal of International Economics, 32, pp. 109-129.

Julius, DeAnne (1990): Global Companies and Public Policy (London: Royal Institute of International Affairs).

Krugman, Paul (1979): "Increasing Returns, Monopolistic Competition, and International Trade," Journal of International Economics, 9, pp. 469-79.

Krugman, Paul (1983): "The "New Theories" of International Trade and the Multinational Enterprise," in D.B. Audretsch and Charles Kindleberger (eds.), The Multinational Corporation in the 1980s, (Cambridge: MIT Press).

Krugman, Paul, and Edward Graham (1992): "The Surge in Foreign Direct Investment in the Late 1980s: Conceptual Issues," forthcoming in Ken Froot (ed.), Foreign Direct Investment Today (NBER).

Lancaster, Kelvin (1980): "Intra-industry Trade under Perfect Monopolistic Competition," Journal of International Economics, 9, pp.151-75.

Markusen, James (1984): "Multinationals, Multi-plant Economies, and the Gains from Trade," Journal of International Economics, 16, p. 205-26.

Rugman, Alan (1988): "The Multinational Enterprise," in Ingo Walter and Tracy Murray (eds.), Handbook of International Management (New York: Wiley).

Swedenborg, Birgitta (1979): "The Multinational Operations of Swedish Firms," (Stockholm: Industriens Utredningsinstitut).

Wilkins, Mira (1970): The Emergence of Multinational Enterprise: American Business Abroad from the Colonial Era to 1914 (Cambridge: Harvard University Press).

Yoffie, David (1992): "Foreign Direct Investment in the Semiconductor Industry," forthcoming in Ken Froot (ed.), Foreign Direct Investment Today (NBER).

</ref_section>