The Multiproduct Firm:
Horizontal and Vertical Integration

by

Elhanan Helpman

Number 335

November 1983

massachusetts institute of technology

50 memorial drive
.cambridge, mass. 02139
The Multiproduct Firm;
Horizontal and Vertical Integration

by

Elhanan Helpman

Number 335

November 1983
The Multiproduct Firm:  
Horizontal and Vertical Integration

by

Elhanan Helpman*
Tel Aviv University and Massachusetts Institute of Technology

1. Introduction

The modern corporation has been shaped in the last hundred and fifty years. Some have argued convincingly that its formation reflects an attempt to efficiently internalize transactions that were traditionally market mediated (Williamson (1979, 1981)). Its development in the form of organizational innovations was linked to improvements in transportation and communication as well as to the invention of mass production techniques (Chandler (1977)). Diversification in production and forward integration into distribution were some of the more important characteristics observed already in the early periods of its development (Chandler (1977), Hannah (1980), Kocka (1980) and Levy-Leboyer (1980)). Many corporations are characterized by horizontal and vertical integration.

It is the purpose of this paper to present a model of a horizontally and vertically integrated firm that captures some essential features of the modern corporation yet is simple enough in order to be conveniently used in various applications, including applications that require a general equilibrium

* This paper is based on ideas that were developed during my work on multinational corporations when I was a Visiting Professor in the Department of Economics at Harvard University. Richard Caves was very helpful in the development of my thinking on the structure of firms.
analysis. The theory of the firm that is developed in the following sections applies to industries which produce differentiated products with economies of scale. It is assumed that there exist inputs like management, marketing and product specific R&D, that can simultaneously serve many product lines which are in some sense similar to each other. These inputs are adapted to be product specific and they generate economies of scope. The firm may also use differentiated intermediate inputs. An important ingredient in the firm's decision process is the choice of the range of products that it can profitably produce, including intermediate inputs. Thus, the product mix is endogeneous to the firm and its choice brings about horizontal and vertical integration. The usefulness of the model is demonstrated by reference to two problems; the regulation of industries and the formation of multinational corporations.

The paper is structured as follows: Section 2 describes a simple form of preferences for differentiated products which is used in order to derive the demand functions faced by monopolistic firms. A model of the horizontally integrated firm is developed in Section 3 by abstracting from the vertical structure of production. The relevance of this model to the regulation of industries is discussed and a small scale general equilibrium model is presented. In Section 4 the model is extended in order to deal with vertical integration. An application of the extended model to the study of multinational corporations and international trade is then described. The closing section provides some concluding comments.
2. Preferences

In order to concentrate on production decisions I simplify the demand side as much as possible. Assume, therefore, that there exist many commodities, one of them, good X, being a differentiated product. There exists a continuum of varieties of product X which can be represented by points on the real line; every point representing a different variety. Moreover, the closer two varieties are to each other on the line, the better substitutes they are for each other in production in a sense to be described below. However, in consumption all varieties are equally well substitutable for each other with the elasticity of substitution \( \sigma \) being constant and larger than one. Formally this can be represented by a utility function \( U(..., u^X) \) which depends on consumption of goods other than X and on the subutility level \( u^X \) derived from the consumption of varieties of the X-product, with

\[
u^X = \int_{\omega \in \Omega} \left[ X(\omega) \right]^{1-\frac{1}{\sigma}} \omega \right\{1- \frac{1}{\sigma}\}^{-1}, \quad \sigma > 1
\]

where \( X(\omega) \) is consumption of variety \( \omega \) and \( \Omega \) is the set of varieties available to consumers. It is well known that for this utility function the demand for variety \( \omega' \) can be represented by:

\[
(1) \quad X(\omega') = \frac{[p(\omega')]^{\sigma}}{\int_{\omega \in \Omega} [p(\omega)]^{1-\sigma} \omega \right\{1- \frac{1}{\sigma}\}^{-1}} \cdot E^X, \quad \text{for all } \omega' \in \Omega
\]

where \( E^X \) is total spending on the differentiated product. In this case the elasticity of demand for a single variety is constant and equal to \( \sigma \) -- the
elasticity of substitution between every pair of varieties. In what follows it is assumed that producers believe that they face the following demand function for every variety of product $X$:

$X(\omega) = k[p(\omega)]^{-\sigma}$ for all $\omega \in \Omega$

where $k$ is constant from the producer's point of view, but it does depend on the aggregate behavior of the industry. Alternatives are discussed in the following section.

This specification of preferences is convenient because it implies the same constant elasticity of demand for every variety of $X$. A possible alternative (which is more appealing) is to use Lancaster's representation of preferences for differentiated products (see Lancaster (1979)), which allows for varying degrees of substitutability in consumption across varieties. However, since the purpose of this paper is to present a new model of production, I have chosen to employ the simplest possible specification of preferences.
3. **Horizontal Integration**

The discussion in this section abstracts from the vertical structure of production and concentrates instead on the horizontal structure. It is assumed for simplicity that there exist only two factors of production; labor $L$ and an input $H$ which will be referred to as the shared input (a generalization to many inputs is not difficult). Varieties of $X$ are produced by means of labor and services of the shared input. A firm that hires the shared input has to adapt it at a cost to serve best a particular variety, which will be referred to as the firm's central variety. Once adapted, the input becomes a firm-specific asset in the sense used by Williamson (1981), and it can also serve the production of other varieties, but with declining efficiency the further away a variety is from the central variety. This loss of efficiency is reflected in the need to employ more labor per unit output. Moreover, the productivity of the services of the shared input may depend on the number of product lines that it has to serve. Thus, borrowing the terminology of local public finance, the shared input can be a 'pure' firm specific public input or it can be a 'congestable' firm specific public input. Common inputs that fit this description are management, distribution, and product specific R&D; they can be adapted to serve best a certain variety of the product and they can simultaneously serve the production and sales of other varieties, but becoming less and less effective in the serving of varieties which are farther away from their central occupation. Management, for example, may be considered to be congestable, in the sense that although the same management team can manage many product lines at the same time its efficiency declines the more product lines it has to manage. The importance of shared inputs is discussed in general terms in Panzar and Willig (1981) and Bailey and Friedlaender (1982), while their importance in the operation of multinational corporations is discussed in Caves (1982, chp. 1).
Consider a firm that hires a quantity $h$ of the shared input and adapts it to its central variety. The costs of hiring the shared input and the adaptation costs are fixed costs at the level of the firm. Apart from these there exist fixed costs which are specific to every product line. Given $h$ it is assumed that in order to produce in the range $(\delta, \delta+d\delta)$ the quantity $x(\omega)$ of varieties which are located at distance $\delta$ (on the real line) from the central variety the required quantity of labor is represented by $\lambda[x(\omega), h, \delta(\omega), m]d\delta$, where $\delta(\omega)$ is the distance of variety $\omega$ from the central variety, $m$ is the measure of the set of varieties produced by the firm and $\lambda(\cdot)$ is the inverse of a quasi-concave increasing returns to scale production function (i.e., increasing returns to scale in $(\lambda, h)$ holding $(\delta, m)$ constant). I will refer to $m$ as the number of varieties produced by the firm. The labor requirement function $\lambda(\cdot)$ is assumed to be increasing in output, decreasing in $h$, convex in $(x, h)$, increasing in $\delta$ and nondecreasing in $m$. The last element describes possible congestability of the shared input.

Total costs of the firm consist of the fixed costs involved in hiring and adapting the shared input plus production costs of the various varieties. Let $C^a(w_L, w_H, h)$ be a cost function, derived from a quasi-concave and increasing returns to scale production function that describes adaptation costs, where $w_L$ is the wage rate and $w_H$ is the reward to a raw unit of the shared input. Let also $\Delta$ be the set of varieties produced by the firm with $x_{\Delta}$ standing for the description of the quantities of all varieties in $\Delta$ that are being produced, and let $m(\Delta)$ be the measure of $\Delta$. Then the firm's cost function is:

$$C[w_L, w_H, x_{\Delta}, m(\Delta)] \equiv \min_{h \geq 0} \left\{ w_H h + C^a(w_L, w_H, h) + w_L \int_{\omega \in \Delta} \lambda[x(\omega), h, \delta(\omega), m(\Delta)] \ d\delta(\omega) \right\}$$
This cost function contains all the above discussed components.

Given the assumption that \( l(\cdot) \) is increasing in \( \delta \) and the fact that the demand function for every variety is the same (see (1')), it is clear that a firm that chooses to produce a subset of varieties in a range in which no other firm produces (as will be the case in the below discussed equilibria) finds it most profitable to produce a connected set \( \Delta \) which is located symmetrically around the central variety. The reason is that for every set of output levels revenue is independent of the varieties that are being produced while costs are lowest when the set of varieties is chosen to be connected (except for a subset of measure zero) and symmetric around the central variety. For this reason we restrict the discussion to sets \( \Delta \) which are connected and symmetric around the central variety. In this case it is appropriate to also call \( m \) the horizontal span of the firm, and the cost function for this type of \( \Delta \) choice can be represented by:

\[
C(w_L, w_H, x_\Delta, m) = \min_{h \geq 0} \left\{ w_H h + C^a(w_L, w_H, h) + 2w_L \int_0^{m/2} l[x(\delta), h, \delta, m] d\delta \right\}
\]

where \( x(\delta) \) stands now for the output level of a variety located at distance \( \delta \) from the central variety. This cost function exhibits ray economies of scale, and for sufficiently small \( m \) also economies of scope (see Willig (1979) for a discussion of these concepts). The ray economies of scale mean that for a given horizontal span a proportional increase in the output level of all varieties within this span increases costs less than proportionately. The economies of scope mean that given \( m \) it is cheaper to produce the output levels \( x_\Delta \) with a single firm (\( h \) adapted to a single variety) than with two or more firms (\( h \)
adapted to two or more varieties). The proofs of these claims for the general cost function defined in (2) are given in the appendix. Given these properties it seems natural to identify the firm with the specialized shared input h. This identification is similar to the commonly used identification of the firm with the input of entrepreneurial ability (see, for example, Oi (1983)).

One may want to ask at this stage whether the product lines in the horizontal span \( \Delta \) have to be part of the firm that owns the specialized resource h. It seems possible to have an organizational structure in which the firm sells the services of h to independent producers of the varieties in \( \Delta \). At this point one applies the organizational argument. Since h is a specialized resource, a producer who chooses to produce a variety in \( \Delta \) and to purchase the services of h will incur fixed costs on the product line and find it difficult to obtain guarantees for the supply of h services on desirable terms. Under these circumstances the production process is rationalized by integration (see Klein et al. (1978)).

The intuition behind the existence of economies of scope for small horizontal spans of the firm can be seen with the aid of Figure 1. In Panel A labor employment per variety is drawn for a firm that produces the same output level \( x \) of all the varieties which are located symmetrically around the central variety \( \omega \) in a connected set. Total labor use in production is represented by the area below this curve. Now suppose that we compare this labor use with the labor use that is required in order to produce the same output levels by two firms, each one using one half of the shared input h and producing half of the varieties. The labor use structure for the two-unit production scheme is
represented in Panel B. Total labor use is represented in Panel B by the area below the two bowled curves. For sufficiently small values of \( m \) the minimum obtained by the curve in Panel B is larger than the minimum obtained by the curve in Panel A (since \( \lambda (\cdot) \) is decreasing in its second argument). The curves from Panels A and B are jointly reproduced in Panel C. The difference in labor use between a two-unit production scheme and a single unit production scheme is measured by the difference between the dotted area and the two striped areas. It is clear from the figure that for sufficiently small horizontal spans this difference is positive, i.e., the dotted area is larger than the striped area, implying the existence of economies of scope. These economies of scope are strengthened when account is taken of adaptation costs which exhibit economies of scale. The point is that for sufficiently small horizontal spans costs are dominated by the size of \( h \) and its inappropriateness is of lesser significance.

The cost structure described by (3) implies that the cost minimizing choice of employment of the firm-specific shared input satisfies:

\[
(4) \quad w_H + \epsilon^h(w_L, w_H, h) = -2w_L \int_0^{m/2} \lambda_h(x(\delta), h, \delta, m) \, d\delta
\]

The left hand side represents marginal costs of the shared input. They consist of marginal (equals average) hiring costs and marginal adaptation costs. These marginal costs should equal the marginal costs saved on account of other factors of production, which in the present context are labor costs, and they are represented by the right hand side. An addition of a marginal unit of the shared input reduces the required labor employment on the line that produces \( x(\delta) \) by

\[
- \lambda_h(x(\delta), h, \delta, m) \, d\delta \equiv - \partial \lambda[x(\delta), h, \delta, m]/\partial h \, d\delta,
\]

which when summed up over all product
lines provides a measure of total labor savings. Hence the interpretation of the right hand side of (4). This is a version of Samuelson’s condition for the optimal supply of public goods applied to public inputs (see Sandmo (1972)).

The cost function described in (3) is concave in factor rewards and its partial derivative with respect to a factor reward equals total employment of the factor of production. Thus, the partial derivative of costs with respect to the shared input equals the quantity of the shared input used in the adaptation process (= \( \delta C^a(w_L,w_H,h)/\delta w_H \)).

It is assumed that the firm engages in profit maximization. This assumption per-se is not enough to determine the product mix and employment choice of the firm. For it is important to know how the firm perceives the reaction of its rivals in the industry to its own strategy. Here there are several choices to be made, depending on the circumstances to which the model is being applied. If the circumstances justify the assumption of monopolistic competition à la Chamberlain, then a typical firm in the industry is assumed to choose its product mix and its pricing strategy; i.e., the pair \((\Delta, p_\Delta)\), where \(p_\Delta\) is a description of the prices of all varieties in \(\Delta\), so as to maximize profits, taking as given the product mix and pricing strategies of its competitors. This seems to be an acceptable characterization for industries with a large number of competing firms, and in this case it seems reasonable to assume that the firm considers \(k\) in the demand function \((1')\) to be unaffected by its choices. If, on the other hand, the number of competitors in the industry is small, it may seem more appropriate to assume that firms play a Cournot game and that they choose \((\Delta, x_\Delta)\). Here a firm may take the product mix and output choice of its rivals as given or it may take the product mix as given and conjecture output responses to its own
choice. In the latter case the demand functions given in (1) can be inverted and
the elasticity of demand facing an individual firm can be calculated for desired
conjectures (see, for example, Spence (1976) and Koenker and Perry (1981)). In
what follows I proceed with the analysis under the assumption that a typical firm
chooses its product mix and pricing strategy \((\Delta, p_\Delta)\) taking as given the product
mix and pricing strategies of its competitors. Moreover, I assume that it
considers \(k\) in the demand equation (1') to be insensitive to its actions. Under
these assumptions the elasticity of demand for a single variety is constant and
equal to the elasticity of substitution \(\sigma\).\(^1\) It is clear that under these
circumstances a firm that seeks to maximize profits choses a product mix that
contains no varieties which are supplied by its rivals, unless it cannot chose a
connected set \(\Delta\) of the desired horizontal span which has this property due to the
product space being 'crowded' by other producers. However, since the set of
potential varieties is the real line, then as long as there exists a finite
number of firms, each one producing a finite number of varieties in a connected
set, our firm can always find a desired finite size product set that does not
overlap with the varieties produced by its rivals. This is the type of
equilibria considered below.

Based on the above described assumptions a typical producer's problem can be
described as the choice of \(m\) and \(p(\delta)\) for \(\delta \in [0, m/2)\) so as to maximize profits:

\[
\pi = \int_0^{m/2} p(\delta)x(\delta)d\delta - C(w_L, w_H, x_{\Delta}, m)
\]

\(^1\) If the firm was to take account of its actions on \(k\) the elasticity of
demand would have been larger and it would have depended on the number of firms
in the industry.
where the cost function is taken from (3) and the right hand side of (1') is substituted for \( x(\delta) \). This is a degenerate problem in the calculus of variations which can be solved by means of pointwise maximization. Assuming an interior solution, one obtains the following first order conditions:

\[
(5) \quad p(\delta) (1 - \frac{1}{\delta}) = w_x l_x \left[ x(\delta), h, \delta, m \right] \quad \text{for all } \delta \in [0, m/2)
\]

\[
(6) \quad p(m/2) x(m/2) = w_x l_x \left[ x(m/2), h, m/2, m \right] + w_m \int_0^{m/2} l_m \left[ x(\delta), h, \delta, m \right] d\delta
\]

where \( l_x (\cdot) \) is the partial derivative of \( l(\cdot) \) with respect to output and \( l_m (\cdot) \) is the partial derivative of \( l(\cdot) \) with respect to the size of the horizontal span. The former represents marginal labor costs of an output expansion while the latter represents marginal costs associated with the loss of efficiency of the shared input that results from an expansion of the product range. Condition (5) is the familiar marginal revenue equals marginal costs condition; it has to be satisfied for every product line. Condition (6) is the condition for an optimal horizontal span. The left hand side represents marginal revenue from the extension of the product range on one side of the central variety. The right hand side represents marginal costs of this extension. There are two components to these marginal costs. The first component represents the costs that have to be incurred on the new product line. The second component represents the additional costs that have to be incurred on existing product lines due to the

---

2 The profit maximization problem is usually not concave. The nonconcavity results from the choice of the horizontal span. It is, therefore, necessary to make sure in particular applications that there exists an interior solution.
fact that the horizontal expansion reduces the efficiency of the shared input on existing product lines as a result of it having to service a new product line. This loss of efficiency can be compensated for by larger labor employment whose value constitutes the second marginal costs component.

Conditions (5) and (6) are explained diagrammatically in Figure 2, which is drawn on the assumption that the shared input is not congestable (i.e., \( l_m(*) = 0 \)) and that marginal labor costs are the same on every product line (i.e., \( l_x(*) = 0 \)), which means that \( l(*) \) has the separable form \( l_1[x(\delta),h] + l_2(h,\delta) \). Given the demand curve \( D \) (from (1')) and the marginal cost curve \( MC \), profit maximization on a product line is achieved at the output level \( x \) at which the marginal revenue curve \( MR \) intersects the marginal cost curve (point A). The corresponding price \( p \) is read off the demand curve at point E. Since every variety has the same marginal revenue and marginal cost curves, it is profitable to charge the same price and to produce the same output level of every variety in the optimal product range. The question that remains is therefore how to determine the product range. For this purpose observe that we can draw in Figure 2 a family of average production costs curves \( w_1(x,h,\delta)/x \) indexed by \( \delta \); the further away a variety is from the central variety the higher its average cost curve. The curves \( AC(\delta_1) \) and \( AC(\delta_2) \) describe two members of this family, with \( \delta_2 > \delta_1 \). It is seen in the figure that at the output level \( x \) average costs associated with \( AC(\delta_1) \) are lower than \( p \) while average costs associated with \( AC(\delta_2) \) are higher than \( p \). It is therefore clear that varieties whose distance from the central variety is equal or larger than \( \delta_2 \) generate a revenue which does not cover costs of production while varieties whose distance from the central variety does not exceed \( \delta_1 \) generate a revenue which is larger than their costs of production. Hence, it is unprofitable to produce the former subset and it is profitable to
produce the latter subset. It is also clear from this argument that the AC curve that passes through point E, drawn as AC(m/2), determines the profit maximizing limit of the horizontal expansion (it is tangent to the demand curve at E as in Chamberlain (1933), except that here it applies only to the extreme variety).

Another way to represent the optimal horizontal span for the case discussed in Figure 2 is by means of the diagram in Figure 3. The central variety is represented by ω on the real line. The line px represents revenue per variety, so that the area below this line represents revenue. The curve \( w_L(\cdot) \) represents costs of production per variety when each one is produced in the same quantity \( x \). Hence, the area below the cost curve represents costs of production. Maximum profits are obtained in the product range \((ω-m/2, ω+m/2)\), where the limiting points are determined by the intersection of the cost curve with the revenue line. The shaded area represents the excess of revenue over production costs. If this difference covers the costs of hiring \( h \) and adapting it to \( ω \) the firm stays in business; otherwise it leaves the industry.

One can use the firm's first order conditions for profit maximization (4)-(6) to analyze the responses of output, employment and the size of the horizontal span to changes in the demand level and its elasticity as well as to changes in factor prices and the technology. Generally speaking these interactions are quite complex and unambiguous predictions are hard to come by without more detailed information about the technology. There are, however, certain technologies for which intuitive responses emerge. Take, for example, the technology that was used to draw Figure 2, and consider the affects of a decline in demand through a fall in \( k \). It is easy to see from Figure 2 that the demand contraction reduces proportionally the demand curve and the marginal
revenue curve, bringing about a decline in output per variety \( x \) and a narrowing of the horizontal span for a given \( h \). However, these two responses induce a reduction in the employment of the shared input (via (4)) which implies an upward shift of the marginal cost curve and the entire family of average cost curves in Figure 2. These cost affects of the reduction in \( h \) reinforce the contraction of output and the horizontal span.

This analysis is of interest. It sheds, for example, a new light on the relationship between entry and the degree of product diversity. The downward shift of the demand curve that was discussed above can result from reduced consumer spending on the industry's product or from entry of new firm's into the industry (see (1)). In the latter case entry is associated with increased product diversity when firms are single product firms (e.g, Spence (1976), Dixit and Stiglitz (1977) and Koenker and Perry (1981)). However, when firms supply many product varieties entry increases the number of firms but reduces the number of varieties produced by a representative firm. This observation is important for the choice of a proper regulatory policy. For regulation of the number of firms in the industry, which is an often considered option (see the above cited references), does not coinside now with the regulation of the number of varieties supplied to consumers. Therefore its relative merit as comapred to, say, output regulation becomes weaker. It also suggests that there is room to consider joint regulation of entry and the horizontal span of firms.

So far we have discussed the behavior of a single firm. It is, however, easy to extend the single-firm analysis to the industry level as long as we consider symmetrical equilibria in which all firms look alike except for their choice of varieties, with no overlaps across firms in variety choices. In
particular, consider a Chamberlinian equilibrium in which entry brings profits down to zero. In such a symmetrical equilibrium (1) and (1') imply:

\[ k = 2^n \int_0^{m/2} \left[ p(\delta) \right]^{1-\sigma} \delta \, dE_X \]

where \( n \) is the number of firms, and the zero profit condition implies:

\[ E_X/n = C(w_L, w_H, x_A, m) \]

i.e., total spending per-firm equals costs per-firm. The equilibrium conditions (4)-(7) enable now an analysis of industry response to changes in the spending level on its product, in factor prices, etc.

As I have pointed out in the introduction the usefulness of this model stems from its relative simplicity which enables one to use it in various applications, and in particular in applications in which general equilibrium analysis is important. For this reason I closed this section with a description of a general equilibrium system in which there exist multiproduct firms of the above described nature. Consider a two sector economy; one sector producing a homogeneous product \( Y \) and the other producing a differentiated product \( X \). The homogeneous product is produced with \( L \) and \( H \) under constant returns to scale, with the associated unit cost function \( c_Y(w_L, w_H) \), while the varieties of the differentiated product are produced under the above described conditions, with the labor requirement function having the separable form:

\[ \ell(x, h, \delta, m) = \ell_1(x, h, m) + \ell_2(h, \delta, m) \]
We have seen that under these conditions profit maximization requires equal pricing of all varieties within the firm's horizontal span and equal output levels of all these varieties. Hence, using (3), I define a cost function restricted to equal output levels by:

\[(8) \quad C_X(w_L, w_H, x, m) = \min_{h \geq 0} \left[ w_h + C^X(w_L, w_H, h) + w_L m^2 \ell_{1}(x, h, m) + 2w_L \int_{0}^{m/2} \ell_2(h, \delta, m) d\delta \right] \]

where \( C_X(\cdot) \) has the usual properties of a cost function with respect to \((w_L, w_H, x)\). Under these circumstances the condition of marginal cost pricing in sector \( Y \), taking the homogeneous product to be the numeraire, and conditions (5)-(7) in the differentiated product industry can be written as follows:

\[(9a) \quad 1 = C_Y(w_L, w_H) \]
\[(9b) \quad p(1 - \frac{1}{\sigma}) = \frac{C_X(w_L, w_H, x, m)}{m} \]
\[(9c) \quad px = C_X(w_L, w_H, x, m) \]
\[(9d) \quad pxm = C_X(w_L, w_H, x, m) \]

Here condition (9a) assures marginal cost pricing of the homogeneous product, (9b) assures that for every variety of the differentiated product marginal revenue equals marginal costs, (9c) assures that marginal revenue from an expansion of product range equals marginal costs of such an expansion, while (9d) assures zero profits for every firm.
Using the properties of the above defined cost functions the factor market clearing conditions can be written as:

\[(9e)\quad a_{LY}(w_L, w_H) y + A_{LX}(w_L, w_H, x, m)n = \overline{L}\]

\[(9f)\quad a_{HY}(w_L, w_H) y + A_{HX}(w_L, w_H, x, m)n = \overline{H}\]

where \(a_{iY}(w_L, w_H) = \frac{\partial c_Y(w_L, w_H)}{\partial w_i}, \quad i=L,H; \quad A_{iX}(w_L, w_H, x, m) = \frac{\partial c_X(w_L, w_H, x, m)}{\partial w_i}, \quad i=L,H; \quad y\) is the output level of the homogeneous product; \(n\) is the number of firms in the differentiated product industry; and \((\overline{L}, \overline{H})\) are the endowments of labor and the shared input. The left hand side of (9e) represents aggregate demand for labor. It consists of labor demanded by the sector producing the homogeneous product, which equals labor demand per-unit output \(a_{LY}(\cdot)\) times the output level \(y\), plus labor demanded by the differentiated product sector, which consists of labor demanded by a representative firm \(A_{LX}(\cdot)\) times the number of firms in the industry. A similar interpretation applies to the shared input market clearing condition (9f). The number of varieties available to consumers equals \(nm\); i.e., the number of firms times the number of varieties produced by a representative firm.

Finally, the system can be closed by a specification of intersectoral preferences. If these are assumed to be of the Cobb-Douglas type, consumers spend fixed budget shares on the homogeneous product and on the differentiated product. If \(s\) stands for the budget share of differentiated products the commodity market clearing condition can be written as:
The system of equations (9a)-(9g) presents a simple general equilibrium system which determines the vector of prices \((p, w_L, w_H)\), output levels \((y, x)\), the number of firms in the differentiated product industry \((n)\) and the horizontal span of a representative firm \((m)\).\(^3\) One can use the methods developed in Jones (1965) in order to obtain comparative statics results for this system, and one can employ various modifications of it in order to study questions of interest. An example would be the problem of regulation of the differentiated product industry that was mentioned above, but this time in a general equilibrium framework (compare, for example, with Horn (1983, chp. 3)). Thus, if the regulator chooses to regulate the number of firms, condition (9d) (i.e., the zero profit condition) need not hold, and instead of it the number of firms is determined by the regulator. Given the number of firms all other endogeneous variables are determined. If, on the other hand, the regulator choses to regulate the output level \(x\), then (9b) (i.e., the condition that marginal revenue equals marginal costs) need not hold, and the other endogeneous variables are determined conditionally on \(x\). The various policies can be compared in welfare terms by means of the following utility indicator:

\[
V = (nm)^{\sigma/(\sigma-1)} y^{1-s} x^s
\]

which results from the preferences presented in Section 2 when the upper level utility function is Cobb-Douglas.

---

\(^3\)System (9) can be adapted to take account of conjectural variations. This is done by replacing \(\sigma\) in (9b) with \(\sigma+e(n)\), where \(e(*)\) is derived from the specified conjectures (as, for example, in Koenker and Perry (1981)). Moreover, a somewhat more complicated system can be developed for Lancaster (1979) type preferences.
4. Vertical Integration

This part of the paper is devoted to an extension of the theory of the firm in order to deal with vertical integration. In what follows the usefulness of the current extension is discussed with reference to multinational corporations which are typically observed to engage in intra-firm trade in intermediate inputs (or middle products). I assume for simplicity that the production of every unit of the final differentiated product requires one unit of the middle product. The middle product is also differentiated and it can be represented by points on the real line. Choose a representation of the middle product which coincides with the representation of the finished good such that a point on the line represents now a variety of the finished good and also the variety of the middle product which is best suited for its production. The meaning of a best suited variety of the middle product is similar to the one postulated for the shared input; the further away the employed variety of the middle product is from the best suited variety, the more labor per unit output has to be employed in the production of the finished good. It is clear that in this case a firm that chooses to produce a connected set of varieties of the final good with a single variety of the middle product will choose the variety of the middle product that is best suited for the production of its central variety.

Intermediate inputs are produced by means of services of the shared input and labor. A firm that produces finished products can use its shared input to service the production of middle products. This is a reasonable assumption for manufactured goods which require similar technologies in the production of components and finally assembled products. This feature of production generates an incentive for vertical integration; an incentive that is strengthened when intermediate inputs are produced under increasing returns to scale (which are
'large' but not 'too large') as is assumed in what follows. For with increasing returns to scale in the production of middle products, which stem partly from fixed costs of positive measure, it is not profitable to provide best varieties of the middle product for all the varieties of the finished good that are being produced. A firm that produces a spectrum of varieties of the finished good will use typically only a small number of varieties of the middle product (here the qualification that the returns to scale are not too strong is important), and if the horizontal span of firms is such that they do not overlap in product space in the sense that no variety is produced by more than one firm (which is the case discussed below), then a duopoly situation may arise between an independent supplier and user of the middle product. This market structure reinforces the rational for vertical integration that was described above, as discussed by Williamson (1971), Porter and Spence (1977) and Klein et al. (1978). This reasoning provides the basis for the formal model of the firm that is developed in this section; i.e., a vertically as well as a horizontally integrated firm. The important point to notice is that this structure of firms stems from an endogenous decision of profit maximizers based on technology and market conditions. For simplicity it is assumed that returns to scale are such that every firm finds it most profitable to produce exactly one variety of the middle product. I will discuss in the sequel the factors that determine the profit maximizing number of middle products.

Suppose that a firm adapts the quantity h of the shared input to its central variety. Suppose also that it uses a single variety of the middle product which is best suited for the production of its central variety. Then it is assumed that in order to produce in the range \((\delta, \delta + \delta)\) a quantity \(x(\delta)\) of a variety of the final product which is located at distance \(\delta\) from the central variety the
required quantity of labor is \(\lambda \int [x(\delta), h, \delta, m] d\delta\), where \(\lambda(*)\) has the properties that were specified in Section 3. In particular, labor requirement rises with the distance from the central variety. This represents the compensation needed for the loss of efficiency of the services of the shared input and in the current specification also the required compensation for the inappropriateness of the middle product, since both are most suitable for the production of the central variety.

Now assume that every variety of the middle product is produced under increasing returns to scale with the associated cost function \(C^Z(w_L, w_H, h, Z)\), where \(Z\) is the output level and \(C^Z(*)\) is declining and convex in \(h\). It is assumed that this cost function contains a fixed cost component which is an 'atom'. Since every unit of the finished good requires a unit of the intermediate input,

\[
Z = \frac{m}{2} \int_0^\infty x(\delta) d\delta
\]

The firm's cost function, which is analogous to (3), can now be defined as:

\[
(3') \quad C(w_L, w_H, x_\Delta, m) = \min_{h \geq 0} \left\{ w_L h + C^a(w_L, w_H, h) + 2w_L \int_0^{m/2} \lambda [x(\delta), h, \delta, m] d\delta + C^a_{Z} w_L, w_H, h, \int_0^{m/2} x(\delta) d\delta] \right\}
\]

It is seen from (3') that the cost minimizing choice of \(h\) satisfies:

\[
(4') \quad w_H + C^a_h(w_L, w_H, h) = -2w_L \int_0^{m/2} \lambda [x(\delta), h, \delta, m] d\delta - C^a_{Z} w_L, w_H, h, \int_0^{m/2} x(\delta) d\delta]
\]
When compared with (4) we see that the marginal costs of hiring and adapting the shared input (the left hand side) are the same in both cases. However, the marginal benefit from an expansion of employment of the shared input includes now an additional element; cost savings in the production of middle products which stem from the fact that the shared input services also the production of middle products. This 'increases' the marginal benefit from expanding employment of the shared input.

Profit maximization leads to a choice of pricing and a horizontal span which bring marginal revenue from each variety in line with marginal costs of output expansion and the marginal revenue from an increase of the horizontal span in line with the marginal costs of this increase. However, in the current case the marginal cost components contain direct plus indirect costs. The counterparts of conditions (5)-(6) are:

\[(5') \quad p(\delta)(1 - \frac{1}{\sigma}) = w_L \int x(\delta),h,\delta,m + \int_{[0,m/2]}^{m/2} C_{12} w_L,w_H,h,2\int_0^\delta x(\delta')d\delta \]' \quad \text{for} \quad \delta \in [0,m/2] \]

\[(6') \quad p(m/2)x(m/2) = w_L \int x(m/2),h,m/2,m + \int_{[0,m/2]}^{m/2} \int_{l,m}^x x(\delta),h,\delta,m \] 
\[\quad + C_{12} w_L,w_H,h,2\int_0^{m/2} x(\delta)d\delta \int x(m/2) \]

The last term on the right hand side of (5') represents the marginal cost associated with output expansion of a particular variety that stems from the need to produce more middle products, while the last component on the right hand side
of \((6')\) represents the marginal cost associated with a one sided expansion of the product range that stems from the need to produce more middle products.

In order to better understand the decision concerning the vertical structure of firms it is helpful to reconsider the assumption that the firm chooses to produce a single variety of the middle product, while in effect it can choose any other number. Suppose for the moment that the firm was to use two varieties of the middle product. In order to calculate the profitability of this case it is necessary to disaggregate \(l(\cdot)\) into the costs that result from the shared input being adapted to the central variety and those that result from the middle product being adapted to some other variety. For simplicity, let

\[
l(x,h,\delta_h,\delta_z,m) = l_1(x,m) + l_h(\delta_h) + l_z(\delta_z)
\]

where \(\delta_h\) stands for the distance of the variety from the central variety and \(\delta_z\) stands for the distance of the middle product from the best suited middle product. Here the additional labor required to compensate for the loss of efficiency of the shared input is \(l_h(\delta_h)\) and \(l_z(\delta_z)\) is the additional labor required to compensate for the loss of efficiency of the intermediate input. I assume that \(l_i(\cdot)\) is an increasing functions, \(i=h,z\). When both the shared input and the middle product are adapted to serve best the central variety, we have \(\delta = \delta_h = \delta_z\). However, with the existence of more types of middle products \(\delta_h\) is generally different from \(\delta_z\). In order to see what is involved consider Figure 4, in which \(\omega_c\) indicates the firm's central variety. Let its horizontal span be given by the end points \(\omega_l\) and \(\omega_r\) with both end points being at an equal distance from the central variety. Then it is clear that due to the assumed cost structure whenever two types of the middle product are being considered it is most efficient to locate the varieties of the middle products at the midpoints between the central variety and the extreme varieties \(\omega_l\) and \(\omega_r\); i.e., at points
ω₁ and ω₂ in Figure 4.

Clearly, all finished goods in the range ω₀ to ωᵣ use the middle product that has been adapted to ω₂ while the others use the middle product that has been adapted to ω₁. In this case profits can be written as:

\[ \pi = 2 \int_{0}^{\frac{m}{2}} p(\delta) x(\delta) d\delta - 2 w_L \int_{0}^{\frac{m}{2}} [x(\delta), h, m] + \lambda_h(\delta)] d\delta \]

\[ - 2 w_L \int_{0}^{\frac{m}{4}} \lambda_z[(m/4) - \delta] d\delta \]

\[ - 2 w_L \int_{\frac{m}{4}}^{\frac{m}{2}} \lambda_z[\delta - (m/4)] d\delta \]

\[ - w_H h - C^a(w_L, w_H, h) - 2 C^z[w_L, w_H, h, \int_{0}^{\frac{m}{2}} x(\delta) d\delta] \]

The first component in the profit expression represents revenue while the second represents direct labor costs exclusive of the labor required to compensate for the fact that the middle products that are being used are not best suited for the varieties of the finished good (except for two points in a continuum). The following two expressions contain these additional direct labor costs. The final three expressions describe the cost of highering the shared input, the cost of adapting it, and the costs of production of the middle product. Using the
demand function (2), the first order conditions for profit maximization over 
$p(\delta); \delta \epsilon [0, m/2)$, h and m are:

$$w_H + C_h^R(w_L, w_h, h) = -2w_L \int_0^{m/2} \lambda_{1h}[x(\delta), h, m] d\delta - 2C_{1h}^R w_L, w_H, h, \int_0^{m/2} x(\delta) d\delta$$

$$p(\delta)(1 - \frac{1}{c}) = w_L \int_{x(m/2)}^{1} \lambda_{1m}[x(\delta), h, m] + \Delta h(m/2) + \Delta z(m/4)]$$

$$p(m/2) x(m/2) = w_L \int_{x(m/2)}^{1} \lambda_{1m}[x(\delta), h, m] - C_{1h}^R w_L, w_H, h, \int_0^{m/2} x(\delta) d\delta] x(m/2)$$

These are substitutes for conditions (4')-(6') which were derived under the 
assumption that the firm produces a single variety of the middle product. Both 
sets of conditions have the same interpretation which needs no repetition. The 
thing to note, though, is the difference in the various marginal costs that 
arises due to the difference in the number of middle products being produced. In 
the condition for optimal employment of the shared input marginal cost savings of 
direct labor use on account of larger use of h is, ceteris paribus, the same in 
both cases. However, marginal cost savings in the production of middle products 
is a multiple of the number of middle products being produced, because the shared 
input serves all product lines. Unfortunately, there is no simple ceteris 
paribus comparison of these cost savings because for given $(w_L, w_H, h, x_A)$ the 
volume of output of middle products is twice as large in the single-variety-
middle-product case as compared to the two-variety-middle-product case. It is 
evertheless clear that if larger use of the shared input reduces fixed costs in 
the production of middle products then the two-variety-middle-product scheme
faces larger marginal cost savings on this account. If, in addition, marginal costs of production of intermediate inputs decline with \( h \) and the decline is larger for smaller output levels, then:

\[
2C_h[w_L, w_H, h, \int_0^{m/2} x(5) d5] > C_h[w_L, w_H, h, \int_0^{m/2} x(5) d5]
\]

and the two-variety scheme can be said to have a marginal advantage in the employment of the shared input as compared to the single-variety scheme. The stronger the affect of the shared input on fixed costs of production of intermediate inputs the more likely this advantage.

A comparison of the pricing conditions of finished products (i.e., marginal revenues equal marginal costs) reveals that, ceteris paribus, two middle products generate lower direct plus indirect marginal costs of production of finished goods if marginal costs of producing intermediate inputs rise with output. Finally, comparing the conditions for the optimal horizontal span two differences emerge. One concerns marginal costs of providing the intermediate input and they are similar to what has been just observed with regard to the pricing conditions. The other concerns costs that arise from a horizontal expansion due to the need to compensate for the inappropriateness of the shared input and the middle product. With the special form of the labor requirement function employed in the derivation of the first order conditions in the two middle products case, these compensations amount to \( l_h(m/2) + l_z(m/2) \) units of labor when a single variety of the middle product is produced and to \( l_h(m/2) + l_z(m/4) \) units of labor when two varieties of the middle product are produced. Hence, ceteris paribus, the firm
faces lower marginal costs of horizontal expansion on this account when it produces two varieties of the middle product, because the quantity of labor required to compensate for the inappropriateness of the middle product in the production of the marginal variety of the finished good is lower in the presence of two types of middle products than it is in the presence of a single type.

It is clear from the discussion of the two middle product case how the analysis generalizes to the case of many varieties of the intermediate input. It is also clear that whether profits are higher when a single variety of the middle product is produced or when two varieties of the middle product are produced depends to a large extent on the size of fixed costs in the production of middle products. If, for example, these fixed costs do not depend on \( h \) and marginal costs do not vary with output, then for high enough fixed costs it is most profitable to produce a single variety of the middle product while if fixed costs are low enough it might be profitable to produce two or more varieties of the middle product. In the limiting case in which these fixed costs are zero it is optimal to produce the entire range of middle products within the domain of the firm's horizontal span.

The model of the horizontally and vertically integrated firm can be embedded in a general equilibrium framework by means of the method described in Section 3. The resulting general equilibrium system is simple enough in order to be used in various applications. One application is discussed in Helpman (1983). There I study conditions for the emergence of multinational corporations and the resulting trade structure. The key assumption in that study is that product lines can be separated from the firm's center and established in different geographical locations without losing excess to the services of the shared input employed in
the center. This enables firms to exploit cross-country differences in factor rewards (as well as other differences). It is shown how horizontally and vertically integrated multinational corporations emerge and how they affect trade patterns. The latter are related to relative country size and cross-country differences in relative factor endowments. There exists intersectoral, intra-industry, and intra-firm trade. The last trade component consists of trade in intermediate inputs and services of the shared input. The vertical structure of production is essential for an analysis of this trade component which has grown in importance in recent decades. The model has proved to be quite useful in this application and it should prove useful in other applications as well.
5. Concluding Comments

The observation that many firms in manufacturing industries are engaged in the production of differential products with increasing returns to scale has brought about the development of theories of monopolistic competition in differentiated products. These theories build on single product firms. Hence, the further observation that modern corporations are horizontally and vertically integrated entities calls for an extension of the theories in order to deal with multiproduct firms. I have proposed in this paper a model of the multiproduct firm in a differentiated product industry which builds on features that have been pointed out in the organizational literature. This model rests on the notion that firms acquire assets (inputs) which are specialized to certain product varieties and which can serve many product lines. These specialized assets bring about horizontal integration and contribute an incentive for vertical integration when they can serve also the production of intermediate inputs. Intermediate inputs are differentiated and they too are specialized inputs. In the process of profit maximization a firm has to choose its product mix, including the intermediate inputs that it will produce for its own use.

This is a rich model of the firm which can be usefully applied to problems in which the nature of firms plays an essential role. I have shown how this model can be embedded in a simple general equilibrium framework and I have discussed its relevance to the study of regulatory policies and multinational corporations. Other applications will undoubtedly also prove useful.
Appendix

I provide in this Appendix proofs of the following properties of the cost function defined in (2):

(a) The cost function exhibits ray economies of scale; i.e.,

$$\lambda \ C[w_L, w_H, x_\Delta, m(\Delta)] > C[w_L, w_H, \lambda x_\Delta, m(\Delta)] \quad \text{for } \lambda > 1.$$  

(b) If $\Delta$ is a connected set and $m(\Delta)$ is sufficiently small, then there exist economies of scope; i.e., for every partition of $\Delta$ into connected subsets $\Delta_i$ ($\Delta = \bigcup_i \Delta_i$):

$$C[w_L, w_H, x_\Delta, m(\Delta)] < \sum_i C[w_L, w_H, x_{\Delta_i}, m(\Delta_i)].$$  

Proof of (a). Let $h^*$ be the solution to the right hand side of (2) given $[w_L, w_H, x_\Delta, m(\Delta)]$. Then, since $C^a(\cdot)$ exhibits nondecreasing returns to scale and $l(\cdot)$ is the inverse of an increasing returns to scale production function in $(l, h)$, we have for $\lambda > 1$:

$$\lambda \ C[w_L, w_H, x_\Delta, m(\Delta)] > w_H \lambda h^* + C^a(w_L, w_H, \lambda h^*)$$

$$+ w_L \int_{\omega \in \Delta} l[\lambda x(\omega), \lambda h^*, \delta(\omega), m(\Delta)] \delta(\omega) \geq C[w_L, w_H, \lambda x_\Delta, m(\Delta)]$$

The last inequality stems from the fact that $\lambda h^*$ need not be the cost minimizing $h$ for output levels $\lambda x_\Delta$.

Proof of (b). Let $h^*_i$ be the solution to the right hand side of (2) given $[w_L, w_H, x_{\Delta_i}, m(\Delta_i)]$. Then due to $C^a(\cdot)$ exhibiting economies of scale and $l(\cdot)$ declining in $h$, we have:
\[ \sum_i C[w_L, w_H, x_{\Delta_i}, m(\Delta_i)] > w_H h^* + C^B(w_L, w_H, h^*) \]

\[ + \sum_i w_L \int_{\omega \in \Delta_i} \lambda[x(\omega), h^*, \delta_i(\omega), m(\Delta_i)] d\delta_i(\omega) \]

where \( h^* = \sum_i h_i^* \) and \( \delta_i(\omega) \) is the distance from variety \( \omega \) to the central variety employed by firm \( i \). Now let \( \delta(\omega) \) be the distance from variety \( \omega \) to some variety in \( \Delta \) which we choose to be the central variety of the single firm. We have the following linear approximation for every \( \Delta_i \):

\[ \lambda[x(\omega), h^*, \delta_i(\omega), m(\Delta_i)] \approx \lambda[x(\omega), h^*, \delta(\omega), m(\Delta)] \]

\[ + \lambda_\delta[x(\omega), h^*, \delta(\omega), m(\Delta)][\delta_i(\omega) - \delta(\omega)] + \lambda_m[x(\omega), h^*, \delta(\omega), m(\Delta)][m(\Delta_i) - m(\Delta)] \]

Clearly, since \( \Delta \) is a connected set, we have:

\[ |\delta_i(\omega) - \delta(\omega)| \leq m(\Delta) \quad \text{for all } \omega \in \Delta \text{ and all } i, \]

and \( |m(\Delta_i) - m(\Delta)| < m(\Delta) \) for all \( i \).

Hence, for \( m(\Delta) \) sufficiently small the last two terms of the approximation become as small as desirable, implying:

\[ \sum_i C[w_L, w_H, x_{\Delta_i}, m(\Delta_i)] > w_H h^* + C^A(w_L, w_H, h^*) \]

\[ + \sum_i w_L \int_{\omega \in \Delta_i} \lambda[x(\omega), h^*, \delta(\omega), m(\Delta)] \geq C[w_L, w_H, x_{\Delta}, m(\Delta)] \]

The last inequality results from the fact that \( h^* \) need not be the cost minimizing employment of the shared input in the single firm.
References


Panel A

Panel B

Panel C

Figure 1
Figure 2

\[ AC(\delta_2) = \omega_L \ell(\cdot) / x \]

\[ AC(m/2) = \omega_L \ell(\cdot) / x \]

\[ D = k \frac{1}{\sigma} - \frac{1}{\sigma} \]

\[ MC = \omega_L \ell(\cdot) \]

\[ MR = (1 - \frac{1}{\sigma}) k \frac{1}{\sigma} - \frac{1}{\sigma} \]
Figure 3
Figure 4