Lectures on International Trade and Investment: Multinational Firms, Offshoring, Outsourcing, and Environment

James R. Markusen
University of Colorado, Boulder
University College Dublin

ETH Zurich May 25-28

Each lecture will be approximately two hours. You will get far more out of the lectures if you work through the background reading ahead of time!

Required background reading is draft chapters (in order of use in lectures) from:

Markusen, James R. and Keith E. Maskus, "International Trade: Theory and Evidence", Wiley publishers, forthcoming early 2011.

Chapter 11: Imperfect Competition and Increasing Returns I: Oligopoly
Chapter 12: Imperfect Competition and Increasing Returns II: Monopolistic-Competition
Chapter 16: Direct Foreign Investment and Multinational Firms
Chapter 17: Fragmentation, Offshoring, and Trade in Services
Chapter 10 Distortions and Externalities as Determinants of Trade

Tuesday, May 25

Lecture 1
Review of the Industrial Organization Approach to Trade
Oligopoly: pro-competitive gains from trade
Monopolistic Competition: gains through increased product and input variety

Lecture 2
Multinational Firms: Preliminaries, Partial-Equilibrium Models
Stylized facts
A general conceptual approach
Partial equilibrium analysis of location: horizontal and vertical firms

Wednesday, May 26

Lecture 3
Multinational Firms: General-equilibrium models, empirical evidence
The horizontal model
The knowledge-capital model
Export-platform strategies
Empirical evidence

Lecture 4
Multinational Firms: Outsourcing versus vertical integration
- Asset specificity, holdup, and incentives
- Physical versus knowledge capital
- The Antrás holdup-model
- The Markusen knowledge-capital and learning model

Thursday, May 27

Lecture 5
Fragmentation, Offshoring, and Trade and Investment in Services
- The general gains-from-trade problem
- The Grossman - Rossi-Hansberg trade in tasks version
- The Markusen - Venables version
- Trade in services and the four modes of supply

Lecture 6
Production and Consumption Distortions and Externalities in General-Equilibrium
- The structure of a general-equilibrium model: the complementarity approach
- Production distortions
- Consumption distortions
- Production externalities

Friday, May 28

Lecture 7
Trade in the presence of Environmental Externalities
- Pollution from production affecting utility
- Pollution from consumption affecting utility
- Production externalities among firms and sectors
- International spillovers

Lecture 8
Trade Policy in the Presence of Environmental Externalities
- Policies and international spillovers
- Cooperative policies and issue linkages
- Tax competition and the race to the bottom

Chapter 11

IMPERFECT COMPETITION AND INCREASING RETURNS I: OLIGOPOLY

11.1 General discussion of increasing returns and non-comparative-advantage gains from trade.

In the previous chapter and also in Chapter 2 (e.g., Figures 2.6 and 2.7), we noted that increasing returns to scale, or scale economies for short, offer the opportunity for gains from trade even for identical economies. Trade and gains from trade can occur without any comparative advantage basis for trade. Trade economists believe that scale economies offer an important explanation for the observation of large volumes of trade between similar economies that we observe in trade data.. In this Chapter and in the next, we will analyze these idea in a more thorough fashion.

In this chapter, we will focus on a case where there are two industries, one producing a good under conditions of perfect competition and constant returns to scale, and a second good under conditions of increasing returns to scale. This IRS good is assumed to be homogeneous, meaning the goods produced by different firms are identical from the consumers' point of view. In the next chapter, we assume that each firm in the IRS sector produces a good that is differentiated from those of its rival firms. In order to avoid some possible confusion, we partly depart from our usual notation here and denote the competitive good as *Y* and the IRS homogeneous good as *X*. This will prove useful in the next section when we have different varieties of *X*.

We will assume that the X firms have technologies in which there is an initial fixed cost of entering production and then a constant marginal cost of added output. We assume a single factor of production *L* (call it labor) which must be divided between the *Y* and *X* sectors and among firms in the *X* sector. Marginal cost is denoted by *mc* and total cost (*tc*) and average cost (*ac*) for an *X* firm are as follows:

$$tc = mcX + fc \qquad ac = \frac{tc}{X} = mc + \frac{fc}{X} \tag{11.1}$$

where *mc* and *fc* are *constants* (parameters), measured in units of labor. Figure 11.1 shows the cost curves of a single firm. It is important to note that average cost is always falling with increased output and it becomes close to marginal cost, but never quite equal to marginal cost. In mathematical terminology, average cost is a rectangular hyperbola, going to infinity as *X* goes to zero and approaching *mc* as *X* goes to infinity.

The consequences of this type of technology for competitive conditions, sometimes referred to as market structure are rather profound. Specifically, this type of increasing-returns technology cannot support perfect competition as a market outcome. This is typically explained by a proof by contradiction. Suppose that there are many small firms such that each firm regards market price as exogenous. If price equaled marginal cost as assumed in competitive theory, then each firm would be making losses since average cost is greater than marginal cost, so this cannot be an equilibrium. Suppose then that the competitive price exceeds marginal cost but again is viewed as a constant by each firm. Then each or any firm knows that once it produces enough output, price will exceed average cost and so profits become positive. But the firm should not stop there, it would maximize profits at the constant price by producing an infinite output and so should expand without bounds. But this cannot be an equilibrium and contradicts the assumption that firms are small. Thus the technology shown in (11.1) and Figure 11.1 is inconsistent with perfect competition.

The equilibrium outcome (assuming an equilibrium exists) must be one in which only a small number of firms are able to survive in equilibrium. Competitive will be imperfect, with each firm having influence over market price. The consequences of the technology in Figure 11.1 are shown for a single monopoly *X* firm in Figure 11.2. Also assume a very simple technology for *Y*, such that one unit of *L* produces exactly one unit of *Y*: $Y = L_y$. *Y* and *L* must have the same price and we will use *L* and therefore *Y* as numeraire, giving them a price of one. $\bar{L}$ is the economy's total endowment of *L*.

The production frontier for the economy, discussed briefly in Chapter 2, runs from $\bar{Y}$ to Y^1 to X^1 in Figure 1 where $\bar{Y}$ to Y^1 is the fixed cost *fc* measure in units of *Y* which is of course also units of *L* in this special case of $Y = L_y$. Suppose that the monopoly equilibrium is at point A in Figure 11.2, where the output of *Y* and *X* are Y^0 and X^0. Then the average cost of producing *X* at point A is given by the total amount of labor needed for *X* divided by the output of *X*.

$$ac = \frac{\bar{L} - L^0}{X^0} = \frac{\bar{Y} - Y^0}{X^0} \tag{11.2}$$

But note that this is the slope of the dashed line in Figure 11.2, the line running from point A to the end of the production frontier $\bar{Y}$. The consequence of this is that, if the monopoly *X* firm is to at least break even, then the price line at equilibrium A must be at least as steep as the dashed line giving average cost. Such a price is shown at p^a in Figure 11.2, where *p* gives the price of *X* in terms of *Y* (and therefore in terms of *L*).

A crucial point is that the market equilibrium involves a market failure: the market outcome is not optimal. The optimum would be a tangency between an indifference curve and the production frontier, implying price equal to marginal cost, but we have shown that this is impossible because it involves the firm making losses. With increasing returns to scale that are "internal" to individual firms as we are assuming here, returns to scale are going to be inherently bound up with imperfect competition. Too little *X* is produced at too high a price.

That is a rather long preliminary that now allows us to get to the heart of the issue: the opportunity to trade is going to offer economies the possibility of pro-competitive gains from trade. Countries will gain from the benefits of increased competition that makes firms produce more at lower average costs (higher productivity) and lower prices. We will focus on putting together two identical economies. Within this setting we consider two cases, one in which the number of firms is fixed before and after trade opens up and one in which there is free entry or exit of firms in response to trading opportunities.

11.2 Pro-competitive gains: the basics

Any discussion of imperfect competition must start with some assumption about how firms react to one another, since each firm is going to be large relative to the market. In this chapter, we are going to assume what is called Cournot-Nash (or Cournot for short) competition in which firms pick a quantity as a best response to their rivals' quantities. Cournot equilibrium occurs when each firm is making a best response to its rivals' outputs. Algebraically, this will be modeled as each firm picking an output quantity holding rivals' outputs constant.

Revenue for a Cournot firm i and selling in country j is given by the price in j times quantity of the firm's sales. Price is a function of all firms' sales.

$$R_{ij} = p_j(X_j)X_{ij}. \qquad \text{where } X_j \text{ is total sales in market j by all firms} \qquad X_j = \sum_i X_{ij} \tag{11.3}$$

Cournot conjectures imply that $\partial X_j/\partial X_{ij} = 1$; that is, a one-unit increase in the firm's own supply is a one-unit increase in market supply. Marginal revenue is then given by the derivative of revenue in (11.3) with respect to firm i's output (sales) in j.

$$\frac{\partial R_{ij}}{\partial X_{ij}} = p_j + X_{ij}\frac{\partial p_j}{\partial X_j}\frac{\partial X_j}{\partial X_{ij}} = p_j + X_{ij}\frac{\partial p_j}{\partial X_j} \qquad \text{since } \frac{\partial X_j}{\partial X_{ij}} = 1 \;\; \text{(Cournot)} \tag{11.4}$$

Now multiple and divide the right-hand equation by total market supply and also by the price.

$$\frac{\partial R_{ij}}{\partial X_{ij}} = p_j + X_{ij}\frac{\partial p_j}{\partial X_j} = p_j + p_j\frac{X_{ij}}{X_j}\left[\frac{X_j}{p_j}\frac{\partial p_j}{\partial X_j}\right] \tag{11.5}$$

The term in square brackets in (11.5) is just the inverse of the price elasticity of demand, defined as the proportional change in market demand in response to a given proportional change in price. This is negative, but to help make the markup formula clearer we will denote minus the elasticity of demand, now a positive number, by the Greek letter $\eta > 0$. We can then write (11.5) as

$$\frac{\partial R_{ij}}{\partial X_{ij}} = p_j\left[1 - \frac{X_{ij}}{X_j}\frac{1}{\eta_j}\right] \qquad \eta_j \equiv -\left[\frac{p_j}{X_j}\frac{\partial X_j}{\partial p_j}\right] \qquad (\textit{elasticity of demand}) \tag{11.6}$$

The term X_{ij}/X_j in (11.6) is just firm i's market share in market j, which we can denote by s_{ij} . Then marginal revenue = marginal cost is given by:

$$mr_{ij} = p\left[1 - \frac{s_{ij}}{\eta_j}\right] = mc_i \tag{11.7}$$

Marginal revenue in Cournot competition turns out to have a fairly simple form as shown in (11.7). . The term s_{ij}/η_j is referred to as the "markup". As you will note, it looks something like the tax formulas we had in the previous chapter and indeed it is possible to think of the monopolist as putting on sort of a tax that raises price above marginal cost. If you refer back to equation (11.4) the markup relates to how much the market price falls when the firm increases its sales. When the demand elasticity is high, there is only a small fall in price when the firm expands output, and so the markup is small.

The role of the firm's market share is more subtle, but crucial to understanding the whole idea of pro-competitive gains. Suppose that we put two absolutely identical countries together in trade, each having just a single X producer. The firms could just continue to do what they were doing in autarky, and prices etc. would be preserved. But each firm will now understand that, if it increases output by one unit, the increase is spread over twice as many consumers and thus the price will fall by only half as much as it would if the monopolist increased supply by one unit in autarky. We sometime say that the firm now *perceives* demand as more elastic or that the *perceived elasticity of demand* η_j/s_{ij} is higher. This is reflected in (11.7) by the fact that s_{ij} goes from one in autarky to one-half when trade opens. The markup up falls, perceived marginal revenue increases, and each firm has an incentive to increase output.

Figures 11.3 and 11.4 work through the consequences of this. The production frontier shown in Figure 11.3 is the production frontier for each of the two identical countries. Point A is the autarky monopoly equilibrium in each country. When the countries are put together in trade, each firm perceives demand as more elastic according to (11.7) and expands output. Equilibrium is restored when the price of X is forced down so that (11.7) is again an equality for both firms with market shares both equal to one half. The new equilibrium must be at a point like T in Figure 11.3.

Note especially that this must constitute a welfare gain for both countries. Output was too low and the price was too high initially in autarky, essentially a distortion like those of the previous chapter. However, this is quite different in that it is an *endogenous* distortion and the opening of trade can be thought of as reducing that distortion. There is a welfare gain from trade (U^a to U^*)for the identical economies through a lower *X* price (p^a to p^*), and higher *X* output (A to T), and more efficient production: the average cost of a unit of *X* is lower, firms move down their average cost curves in Figure 11.1.

Figure 11.4 shown a somewhat different and equally important case. Suppose that there might be more than one firm in each country initially, and that firms can enter or exit until profit are zero.[1] Then, referring back to equation (11.2) and Figure 11.2, the equilibrium price must be equal to average cost, given by the line connecting the production point in Figure 11.4 with the endpoint of the production frontier $\bar{Y}$. Let point A in Figure 11.4 be the initial autarky equilibrium for each country and let p^a denote the autarky price: the price equals average cost reflecting the entry or exit of firms until profits are zero. The distance $\bar{Y}Y^0$ in Figure 11.4 is now the combined or sum of the fixed costs of all firm active in the market in equilibrium.

Now put the two identical economies together in free trade as before. Each firm individually has an incentive to expand output: equation (11.7) continues to apply to pricing decisions. However, when they all do this, profits will be driven negative. This will cause the exit of some firms in each country until zero profits are re-established. This is represented by the outward shift of the linear segment of the production frontier in Figure 11.4 from $\bar{Y}Y^0$ to $\bar{Y}Y^1$: resources which were being used (uselessly) in fixed costs are now freed up for actual production. The new equilibrium will be at a point such at T in Figure 11.4. Note that it is possible now for the consumption of both *Y* and *X* to increase due to the efficiency gains of a larger world output being produced by fewer firms than the total number in the two countries in autarky. Welfare in each country rises from U^a to U^*.

A typical intelligent question at this point is how firm exit is consistent with more competition? The answer is that there is some exit in each country individually, but more left in free trade in total between the two countries than were in each individual country in autarky. For example, suppose that each country has four firms in autarky. Suppose that trade forces the exit of one firm in each country, leaving three in each country. But that means that there are now six firms in total competing for the business of each consumer instead of four in autarky. Exit in each country individually is quite compatible with increased competition.

The nice outcomes (for free trade advocates at least) shown in Figures 11.3 and 11.4 have not been rigorously established, and then are a number of problems. First, the elasticity of demand which we denoted by η is generally not a constant: it will change as prices and total sales change. While the basic message is clear and some readers and professors will wish to move on, we now proceed to offer two special cases that have been widely used in the literature to more rigorously solve for the effects of trade.

11.3 Special case I: quasi-linear preferences

The first special (or rather specific) case has been widely used in industrial organization as well as in international trade. In the latter, it has been widely used to analyze what is known as “strategic trade policy”, a topic treated later in the book. It will also be used later in analyzing multinational firms.

We can keep the notation simple at first by assuming just a single monopoly firm in each country, so n = 1, and by normalizing the population *L* to the number one, so *X* give both total output of *X*, the output per firm, and the consumption per capita. Assume that the preferences of a representative consumer in each country are given by:

$$U = \alpha X - (\beta/2)X^2 + Y \tag{11.8}$$

The crucial property of these preferences are constant marginal utility for good *Y* and the importance of this will become clear in a minute. This linearity in *Y* has led some authors to call these preferences "quasi-linear". As above, one unit of factor *L* is need to produce one unit of good *Y*, and *p* denotes the price of *X* in terms of *Y* or *L*. Let Π denote profits of the firm and *L* (equal to one initially) the number of workers/consumers. Profits are viewed as exogenous by individual consumers. Then the budget constraint for the representative consumer is given by income (also measured in terms of *L*) equals expenditure.

$$L + \Pi = pX + Y \tag{11.9}$$

Substituting from the budget constraint for good *Y* in (11.8), we have the consumer's choice problem, where profits (Π_i) are viewed as exogenous by an individual consumer.

$$Max(X)\ U_i = \alpha X - (\beta/2)X^2 + L + \Pi - pX \tag{11.10}$$

The (inverse) demand function is given by the first-order condition, the derivative of (11.10) with respect to *X*, and is linear in *X*.

$$\frac{dU}{dX} = \alpha - \beta X - p = 0 \qquad => \quad p = \alpha - \beta X \tag{11.11}$$

The feature of quasi-linear preferences that makes them so attractive is that demand does not depend on income and hence it does not depend on profits. This makes the model much easier to solve. On the other hand, the unattractive feature of these preferences is that the demand for *X* does not depend on income, there is a zero income-elasticity of demand for *X*, surely a totally unrealistic assumption. All added income at a fixed price for *X* will be spent on *Y*.

Consider autarky first. Let Π denote the profits of the domestic firm. These profits are revenues minus marginal costs (*mc* will be denote by just *c*) and minus fixed costs, denoted simply as *f*. Substituting in the demand function in (11.11) for *p*, we get

$$\Pi = pX - cX - f = [\alpha - \beta X]X - cX - f \tag{11.12}$$

The first-order condition for profit maximization gives:

$$\frac{d\Pi}{dX} = \alpha - 2\beta X - c = 0 \qquad => \quad X = \frac{\alpha - c}{2\beta} \tag{11.13}$$

Now assume that two identical economies trade freely. Let the two countries and their firms be denoted with subscripts i and j. There are now twice as many consumers in the integrated world economy and the demand price depends on *per capita* consumption not on *total* world consumption. That is, world price will be unchanged if there is twice as much output in the integrated world as there was in each country individually in autarky. We now have $L = 2$ and so the world price of X will

$$p = \alpha - \beta(X_i + X_j)/L = \alpha - \beta(X_i + X_j)/2 \tag{11.14}$$

Profits for firm i are now given by

$$\Pi_i = [\alpha - \beta(X_i + X_j)/2]X_i - cX_i - f \tag{11.15}$$

Assume Cournot competition, so each firm makes a best response to the other firm's output, maximizing profits holding the other firm's output fixed. The first-order condition for profit maximization is given by

$$\frac{d\Pi_i}{dX_i} = \alpha - \beta X_i - \beta X_j/2 - c = 0 \tag{11.16}$$

There is a corresponding equation for firm j, giving two equations in two unknowns. But since we have assumed that the countries are absolutely identical, we know that the solution will be symmetric with both firms producing the same amount in equilibrium. Exploiting this symmetry, we can solve (11.16) for the Cournot output of the firm i (equal to the country j firm's output) by setting $X_i = X_j$.

$$X_i^* = X_j^* = 2\frac{(\alpha - c)}{3\beta} > \frac{\alpha - c}{2\beta} = X^a \tag{11.17}$$

where the right-hand inequality is the autarky output of each firm given in (11.13). Output of each firm expands by one-third ((2/3)/(1/2) = 1/3) when trade opens. This is the effect shown in Figure 11.3 and it must imply mutual gains from trade for the two countries.

Now let's turn to the case of free entry and exit. Again, let's just think about one market and think of trade between two countries as doubling the size of the market. This allows us to simplify the notation.

Assume that there are L individuals (which can again be normalized to one in autarky as we did above). We will also now need to keep track of the difference between output per firm and aggregate X output. Summing over the number of firms n, per capita consumption is

$$\sum_i^n X_i/L \tag{11.18}$$

The number of firms, n, is now *endogenous*.

Demand and profits for the ith firm are then given by

$$p = \alpha - \beta\left[\sum_j X_j/L\right] \tag{11.19}$$

$$\Pi_i = p_i X_i - cX_i - f = \left[\alpha - \beta\left[\sum_j X_j/L\right]\right]X_i - cX_i - f \tag{11.20}$$

The firm's first-order condition is the derivative of (11.20) with respect to X_i holding all of the other firm's outputs constant. Marginal revenue minus marginal cost for firm i is given by:

$$MR - MC = \alpha - 2(\beta/L)X_i - (\beta/L)\sum_{j\neq i} X_j - c = 0 \tag{11.21}$$

Now we again know that there will be symmetry in equilibrium: the output of any active firm will be the same as that of every other firm. *X* will denote the output of an *individual* (not aggregate) "representative firm", and *n* the number of firms. All firms that are active in equilibrium will produce the same amount.

$$MR - MC = \alpha - (\beta/L)(n + 1)X - c = 0 \tag{11.22}$$

The second equation we need for equilibrium is the free-entry condition that will be associated with the number of firms. Then the zero profit condition is that the profits of the representative firm are exactly zero.

$$\alpha X - (\beta/L)nX^2 - cX - f = 0 \tag{11.23}$$

Multiple (11.22) through by *X*.

$$\alpha X - (\beta/L)(n + 1)X^2 - cX = 0 \tag{11.24}$$

We then have two equations (11.23 and 11.24) in two unknowns, *n* and *X*. Solving these two equations in two unknowns we get

$$X = \left[\frac{Lf}{\beta}\right]^{1/2} \tag{11.25}$$

So output per firm increases with the size of the economy (*L*). But with price equal to average cost, this must also mean that the equilibrium price of *X* falls. Finally, putting (11.25) into (11.22), we can solve for n

$$n = (\alpha - c)\left[\frac{L}{\beta f}\right]^{1/2} - 1 \tag{11.26}$$

The number of firms increases with the square root of the size of the world economy. A restriction that the economies are sufficiently big such that $n > 1$ in autarky is sufficient to imply that when *L* doubles the number of firms less than doubles. But this in turn means that each country individually must have exit relative to the number of firms in autarky. It is much the same as our numerical example in section 11.2 above: each country has some exit yet the integrated world economy has more firms in total than the individual countries did in autarky. The situation is exactly like that shown in Figure 11.4.

In summary then, the combination of increasing returns to scale with imperfect competition means that there are gains from trade even for identical economies under the assumptions used here, regardless of whether or not there is a fixed number of firms (same in each country) or there is free entry and exit in response to the opening of trade. Note finally from Figures 11.3 and 11.4 that these gains from trade for identical economies are not associated with any net trade between the economies. Both countries are at point T in free trade. Yet (given our assumption of costless trade) there could be a lot of two-way gross trade flows, with some consumers arbitrarily buying *X* from the producer in the other country. This is referred to a intra-industry trade and the there is a lot of such trade among the high-income developed countries.

11.4 Special case II: Cobb-Douglas preferences (can be skipped without loss of continuity)

One of the big limitations of the quasi-linear case is that it imposes the assumption that there is a zero income elasticity of demand for *X*. But surely a lot of the manufacturing and service industries that are characterized by increasing returns to scale and imperfect competition are producing goods with high income elasticities of demand. Let *X* again denote the output of an *individual X firm* and assume that there is a fixed number *n* of such firms. *Y* denotes the total output of *Y* as before. Suppose that preferences are Cobb-Douglas and given by

$$U = (nX)^{\alpha} Y^{1-\alpha} \qquad \text{with income } (I) \text{ constraint} \qquad I = \bar{L} + \Pi = pnX + Y \tag{11.27}$$

We treated this exact case earlier in Chapter 2. Continue to let the price of *Y* be numeraire, equal to one. We showed in Chapter 2 that the demand functions are

$$nX = \frac{\alpha I}{p} \qquad Y = (1-\alpha)I \tag{11.28}$$

In this case, if we compute the elasticity of demand for *X*, we will find that $\eta = 1$. This is going to greatly simplify our analysis, though note that we will need to be sure that there is more than one firm producing in each country in equilibrium; that is, the market share of a firm must be less than one in autarky in order for marginal revenue given by (11.7) to be positive. Subject to this restriction, the markup is just the firm's market share, which in turn is just $1/n$. Marginal revenue equals marginal cost is given quite simply by

$$p(1 - 1/n) = c \qquad p = \frac{n}{n-1}c \tag{11.29}$$

So far, so good. But things get messy because income now matters for the demand for *X* and profits are part of income. This is exactly the mess that quasi-linear preferences avoids! Since *X* refer to the output of a single firm and, assuming all firms producing in equilibrium are identical, then *nX* gives total *X* output. Substituting the budget constraint from (11.27) into the demand function (11.28) and writing out the expression for profits, the aggregate demand for *X* is given by:

$$nX = \alpha(\bar{L} + \Pi)/p = \alpha(\bar{L} + n(pX - cX - f))/p \tag{11.30}$$

Rearranging the equation and making use of the second equation in (11.29) to eliminate the endogenous variable *p* gives us

$$(1-\alpha)npX + \alpha n(c)X = (1-\alpha)\frac{n^2}{n-1}(c)X + \alpha n(c)X = \alpha(\bar{L} - nf) \tag{11.31}$$

Dividing through by n(c) and multiplying one term by $(n-1)/(n-1)$, we have

$$\left[(1-\alpha)\frac{n}{n-1} + \alpha\frac{n-1}{n-1}\right]X = \frac{\alpha(\bar{L}-nf)}{n(c)} \tag{11.32}$$

and then an explicit solution for X in the case when *n* is fixed.

$$X = \left[\frac{\alpha(\bar{L}-nf)}{n(c)}\right]\left(\frac{n-1}{n-\alpha}\right) \tag{11.33}$$

Suppose now that we put two identical economies together, each with a fixed number of firms *n*. $\bar{L}$ doubles in (11.33) and so does *n* relative to autarky. Then the first bracketed term on the right-hand side of (11.33) does not change: both the numerator and denominator double. But the second bracketed term increases given $\alpha < 1$. For example, let $n = 2$ in autarky and let $\alpha = 0.5$. Then the introduction of trade ($n = 2$ to n = *4*) increases the second term from 2/3 to 6/7, an increase in output per firm of 29 percent. Once again, this is exactly the situation shown in Figure 11.3. Note from (11.29) that the price of *X* also falls as shown in the Figure.

Now let us consider the free entry and exit version of the Cobb-Douglas case. The marginal revenue, marginal cost equation is unchanged.

$$p(1-1/n) = c \tag{11.34}$$

The free entry or zero-profit equation is given by

$$pX = cX + f \tag{11.35}$$

and, with no profit income, the demand for *X* is given by

$$nX = \alpha\bar{L}/p \tag{11.36}$$

Multiple (11.34) through by *X* and then divide (11.35) by (11.34)

$$\frac{n}{n-1} = 1 + \frac{f}{cX} \quad \Rightarrow \quad \frac{n}{n-1} - \frac{n-1}{n-1} = \frac{f}{cX} \tag{11.37}$$

which gives us output per firm.

$$X = (n-1)\frac{f}{c} \tag{11.38}$$

Multiple both sides of (11.34) by *X*, and substitute for *pX* from (11.36)..

$$p(1-\frac{1}{n})X = p\left(\frac{n-1}{n}\right)X = \left(\frac{n-1}{n}\right)\frac{\alpha\bar{L}}{n} = cX \tag{11.39}$$

Now substitute the expression for *X* in (11.38) to give us a solution for *n*, the endogenous number of firms.

$$\left(\frac{n-1}{n}\right)\frac{\alpha\bar{L}}{n} = c(n-1)\frac{f}{c} \qquad n^2 = \frac{\alpha\bar{L}}{f} \tag{11.40}$$

Take the square root of the right-hand equation to get n and then substitute this into (11.39) to get *X*.

$$n = \sqrt{\frac{\alpha\bar{L}}{f}} \qquad X = \left[\sqrt{\frac{\alpha\bar{L}}{f}} - 1\right]\frac{f}{c} \tag{11.41}$$

With free entry and exit, both the number of firms and the output per firm increase with the square root of the size of the market. Once again, thinking of trade as a doubling of market size, this means that there is some exit in each country individually with the opening of trade. The effects of trade for two identical economies are exactly those shown in Figure 11.4.

11.5 Summary

To this point in the book, we have concentrated on determinants of trade that involve differences between countries. Gains from trade involve exploiting these differences, such as producing and exporting according to comparative advantage. Now we come to a situation in which there can exist gains from trade even between identical economies and indeed we concentrate here on precisely this case. We did touch on the role of increasing returns to scale in the previous chapter, but now we turn to a fuller analysis and assume that scale economies occur at the level of the individual firm, sometimes termed "internal" economies of scale to distinguish them from the external economies of the previous chapter.

The situation quickly becomes complicated because internal or firm-level economies of scale are not compatible with perfect competition and hence the simple tools needed to analyze competitive models need to be extended. Increasing returns to scale and imperfect competition are inherently related to one another. This require introducing new methods and new tools to take into consideration firms with market power and the strategic interaction between such firms. In doing so, we simplified our theoretical economies in other ways, in particular on the factor market side, assuming only a single factor as in the Ricardian model of trade.

In this chapter, we focused on an industry in which firms produce a homogeneous good or alternatively the goods of the different firms are perfect substitutes. Firms compete according to the Cournot-Nash model, choosing quantities that are best responses to the outputs of other firms. Autarky equilibrium for a country is distorted, and we showed that this distortion looks a lot like a production tax as modeled in the previous chapter. Price exceeds marginal cost in market equilibrium, and too little is produced at too high a price to maximize social welfare.

We then open up two identical economies to trade and show that this generates a pro-competitive effect. In technical terms, firms perceive demand as more elastic and hence expand outputs in response to the opening of trade. But when all firms do this, industry output expands, the price and markup falls, average cost falls (or productivity improves) and social welfare increases.

We considered two versions of the model, one in which the number of firms is fixed before and after the opening of trade and one in which there is free entry and exit of firms in response to trade. The effects just mentioned in the previous paragraph are present in both cases, but the free entry/exit case is particularly interesting in that it may be possible for the identical economies to end up consuming more of both goods, the competitive good as well as the increasing-returns good. The important lesson is that trade does not offer welfare gains just based on difference between countries, it also offers gains to very similar countries in terms of more efficient production, lower prices, and high consumption quantities.

REFERENCES

Brander, James A. (1981), "Intra-industry trade in identical commodities", *Journal of International Economics* 11, 1-14.

Helpman, Elhanan. and Paul A. Krugman (1985). *Market Structure and Foreign Trade*, Cambridge: MIT press.

Horstmann, Ignatius J. and James R. Markusen (1986), "Up the average cost curve: inefficient entry and the new protectionism", *Journal of International Economics* 20, 225-228.

Markusen, James R. (1981), "Trade and the gains from trade with imperfect competition", Journal of International Economics 11, 531-551.

Melitz, Mark J. and Gianmarco I.P. Ottaviano (2008), "Market size, trade, and productivity", *Review of Economic Studies* 75, 295-298.

Venables, Anthony J. (1985), "Trade and trade-policy with imperfect competition - the case of indentical products and free entry", *Journal of International Economics* 19, 1-19.

Endnotes

1. The assumption that firms enter or exit until profits are exactly zero means that we are allowing the number of firms to be a continuous variable and not restricted to integer values. This can be puzzling at first, but it is a common trick in economic theory. It allows the problem to be formulated as equations rather than as a difficult (or impossible) to solve integer programming problem.

Figure 11.1

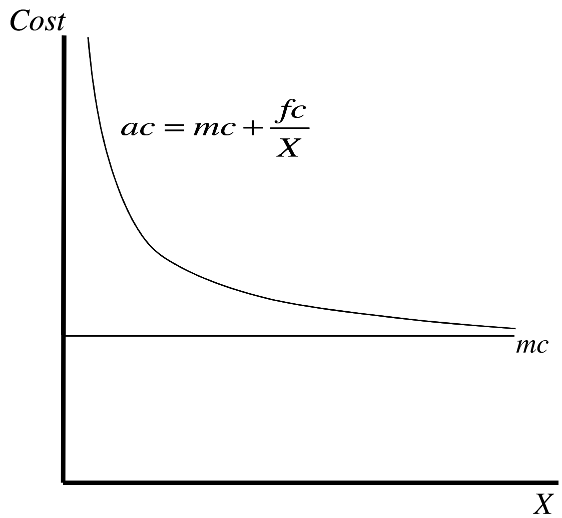


Figure 11.2

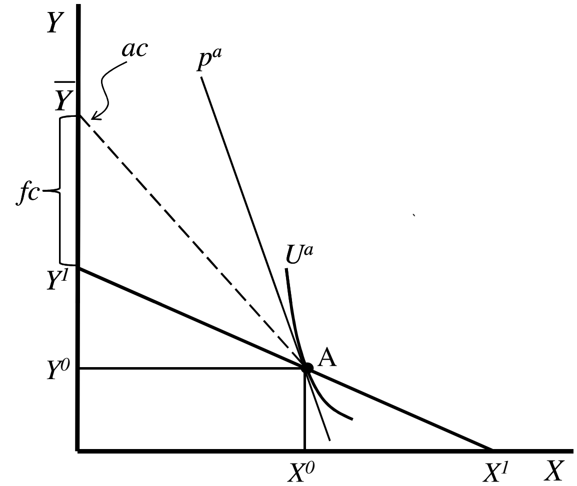


Figure 11.3

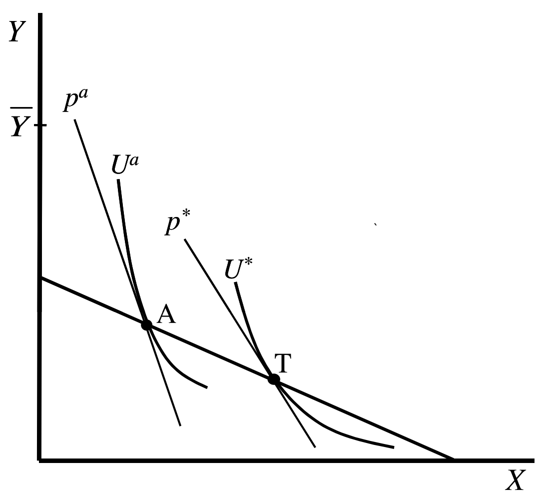
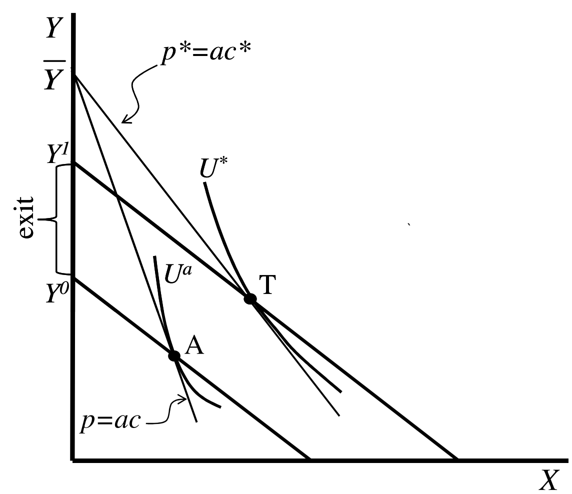


Figure 11.4



Chapter 12

INCREASING RETURNS AND IMPERFECT COMPETITION II: MONOPOLISTIC COMPETITION

9.1 Trade and gains from trade through increased product diversity

The previous chapter introduced economies of scale and imperfect competition as a determinant of trade and gains from trade. We concentrated on a homogeneous-good industry (firms produce identical products) and the idea that scale economies limit the number of firms in the market. This limitation leads to imperfect competition in equilibrium, with products marked up above marginal costs. Firms interact strategically with one another and the consequence of this is that the effective enlargement of the market following the opening of trade leads to pro-competitive gains from trade. Firms move down their average cost curves and consumers benefit from lower prices for the same goods.

In this Chapter, we are again going to look at increasing return to scale and imperfect competition but from a different point of view. We are going to assume that firms produce differentiated products but that the market can support a relatively large number of firms such that there is minimal strategic interaction among the firms. Instead of trade resulting in pro-competitive gains, the same products at lower prices, trade will result in a greater variety of products at the same prices and this will raise consumers' welfare.

These assumptions, that products are different but that there are many firms and minimal strategic interaction, are typically referred to as monopolistic competition. Monopoly refers to the fact that each firm produces a somewhat different product and hence will have influence over its market price even if there are literally hundreds of firms. Competition refers to the fact that there are sufficiently many firms such that there is little strategic interdependence among them.

The following preferences (utility function) are typically used to introduce monopolistic competition and are generally referred to as Dixit-Stiglitz (1977) preferences. This approach is also sometimes referred to as "love of variety" for a reason that should become clear. Let X_i refer to one good in a set of n goods. Ignoring any other sectors for the moment, utility is given by

$$U = \left[\sum_i^n X_i^\alpha \right]^{\frac{1}{\alpha}} \qquad 0 < \alpha < 1 \qquad \frac{1}{\alpha} > 1 \qquad \sigma = \frac{1}{1-\alpha} > 1 \tag{12.1}$$

Many of you will recognize this as a CES function that we introduced earlier in the book, where σ gives the elasticity of substitution among the varieties. It is in fact a special case in which all goods carry the same weighting in producing utility, and in which the elasticity of substitution is restricted to values greater than one. The latter assumption means that indifference curves intersect the axis and so positive utility can be derived from a subset of goods. Indeed, the whole point of this approach is that only a subset of goods gets produced in equilibrium and that the number of goods available is endogenous.

The individual goods in (12.1) are said to be symmetric but imperfect substitutes. Symmetric means that they all have the same weight in producing utility as just noted and hence a consumer is indifferent between one apple and one orange. However, they are also imperfect substitutes meaning that variety is valuable: a consumer would rather have one apple and one orange than either two apples or two oranges. To see this last point, assume that each good that is produced is produced in the same amount, so the that the summation in (12.1) is just n, the number of goods, time X, which we will term the quantity of a "representative good".

$$U = \left[nX^{\alpha}\right]^{\frac{1}{\alpha}} = n^{\frac{1}{\alpha}}X \tag{12.2}$$

We see from (12.2) that utility has constant returns in the amount of each good consumer: double each X and utility doubles. But we also see that utility has increasing returns to scale in the range of product variety (henceforth just "variety"). Let n^0 and X^0 denote the number of goods and the representative quantity produced initially, and U^0 the initial level of utility. Then suppose that the number of goods doubles to $2n^0$ but the quantity consumed of each falls in half to $X^0/2$. Utility is then given by

$$U = (2n^0)^{\frac{1}{\alpha}}(X^0/2) = 2^{\frac{1}{\alpha}-1}(n^0)^{\frac{1}{\alpha}}X^0 = 2^{\frac{1-\alpha}{\alpha}}\left[(n^0)^{\frac{1}{\alpha}}X^0\right] = 2^{\frac{1-\alpha}{\alpha}}U^0 > U^0 \tag{12.3}$$

Equation (12.3) shown that welfare improves when the consumer has half as much of each of twice as many goods. Hence the term "love of variety".

Some intuition is provided in Figures 12.1 and 12.2, where we assume that each of two varieties are produced with increasing returns: a fixed cost and a constant marginal cost as in the previous Chapter. Consider first Figure 12.1, where the production frontier $\bar{X}_2 X_2' X_1' \bar{X}_1$. It may be that in autarky, a country may wish have variety even though it is expensive in terms of having to pay the fixed costs for both goods, and hence prefers the autarky outcome shown as $X^a = D^a$ in Figure 12.1. Trade with the second identical country can then allow each country to specialize in one of the goods, trading half of its output for half of the output of the other country's good. Then both countries can consumer at point D* in Figure 12.1.

Figure 12.1

On the other hand, the high fixed costs and the sacrifice of scale economies may mean that it is better to produce and consumer just one good in autarky, which is the situation shown in Figure 12.2. Here the country achieves utility level U^a in autarky by producing either X_1 or X_2 but not both. This is a higher level of welfare than producing both goods: the added variety is not worth the sacrifice of quantity. Now let the two identical countries get together, with each specializing in one of the goods. Now they can trade to the common consumption point D^* in Figure 12.2, achieving a gain from trade.

Figure 12.2

Note the difference in the source of gains from trade between Figure 12.1 and 12.2. In the former case, the consumer gets more of the same goods, the source of gains in the previous Chapter. In the case of Figure 12.2, the consumer actually gets less quantity of a given good, but enjoys more variety. This is in fact exactly the outcome explored in equation (12.3). It is crucial to note in Figure 12.2 that trade does not result in increased output of any good that is produced initially. No firm moves down it average cost curve, there is no increase in firm scale. Nevertheless it is indeed scale economies that are responsible for the welfare gains in that scale economies limit the number of goods produced initially.

To press this last point a little further, note that if there were no fixed costs and all goods were produced with constant returns to scale, that the optimum under Dixit-Stiglitz preferences would be to have infinitely many goods produced in infinitely small quantities. This is just a logical extension of the argument behind equation (12.2). Increasing returns to scale make diversity costly and hence limit the range of goods in equilibrium.

Figures 12.1 and 12.2 illustrate what can happen, but they by no means prove that this is what will happen in a market environment characterized by imperfect competition.

12.2 A more formal approach to Dixit-Stiglitz and love of variety

We will assume that there are two sectors: sector *X* is composited of firms producing differentiated goods as above, and sector *Y* produces a homogeneous good with constant returns to scale. We will assume a very simple factor market structure much the same as in the previous chapter. There is only one factor of production which we will call labor, and we will use this as numeraire assigning a value of one to the wage rate, $w = 1$. We will assume that the consumer has Cobb-Douglas preferences between Y and X, and CES preferences among the *X* varieties. Much of what follows is based in Krugman (1979, 1981) and Helpman and Krugman (1985).

Total income is given by *L* when the wage is chosen as numeraire. We are also going to assume that all potential *X* varieties have the same cost function. This is a common assumption that, when combined with symmetry in demand, gives us the result that any good that is produced is produced in the same amount and sells for the same price. Henceforth *X* and p_x will denote the price of a representative good which are the same for all goods actually produced. The utility function and the budget constraint for the economy are given by:

$$U = \left[\sum_i X_i^{\alpha}\right]^{\frac{\beta}{\alpha}} Y^{1-\beta} \qquad \sigma = \frac{1}{1-\alpha} \qquad L = np_x X + p_y Y \tag{12.4}$$

If you solve the optimization problem, the consumer's demands for *X* varieties and *Y* are

$$Y = (1-\beta)\frac{L}{p_y} \qquad X_i = p_{xi}^{-\sigma}\left[\sum_i p_{xi}^{1-\sigma}\right]^{-1}\beta L \qquad nX = \beta\frac{L}{p_x} \tag{12.5}$$

The demand response for a given variety in response to a change in its own price is a bit complex, since the variety's own price appears both as the first term on the right-hand side of the second equation of (12.5) but also appears in the summation term inside the square brackets. Thus the derivative of the demand for *X* with respect to its own prices must be found by using the differentiation of a product rule. However, it can be shown (and we will not do so here) that the effect of a change in a firm's price on the summation term in square brackets become extremely small as the number of varieties (firms) n becomes large. As a consequence, most work in this area assumes that an individual firm is too small to affect the summation term in (12.5), an assumption known as "large-group monopolistic competition. Assuming that the term in brackets in (12.5) is viewed as a constant by an individual firm, the price elasticity of demand for an individual goods is given simply by σ, the elasticity of substitution among the *X* goods. Referring back to our derivation of marginal revenue in the previous chapter, the markup takes on the very simple formula 1/σ. The elasticity of demand and marginal revenue are then given as follows.

$$-\frac{p_x}{X}\frac{\partial X}{\partial p_x} = \sigma \qquad mr_x = p_x(1 - 1/\sigma) = mc_x \tag{12.6}$$

Turning to production, marginal cost for *Y*, marginal cost for *X*, and fixed costs of an *X* variety are denoted by mc_y, mc_x, and fc_x respectively. The full general equilibrium model for a single economy is given by a set of inequalities with associated variables as described back in Chapter 4. First, there is are pricing equations for the *Y* industry and for each *X* variety. Second, there is a zero profit condition for each *X* variety, which is typically written as markup revenues equal fixed costs instead of the longer equation for revenues equal total costs. It is useful to think of fixed costs as a produced good, such as factor and equipment, hence there is a pricing equation for factories (fixed costs). These three pricing inequalities are given as follows:

Inequality	Definition	Complementary Variable

$p_y \le mc_y$	pricing for Y	Y	(12.7)
$p_x(1 - 1/\sigma) \le mc_x$	pricing for X	X	(12.8)
$(p_x/\sigma)X \le fc_x$	pricing for n (free entry)	n	(12.9)

Then there are three market-clearing conditions, which require that supply equal demand (strictly speaking supply is greater than or equal to demand). There is demand and supply for Y, for total X production, and for labor. These equations follow from (12.5) and are as follows.

$(1 - \beta)L/p_y \le Y$	demand/supply Y	p_y	(12.10)
$\beta L/p_x \le nX$	demand/supply X varieties	p_x	(12.11)
$(mc_y)Y + n(mc_x)X + n(fc_x) = L$	demand/supply L	w	(12.12)

This model can be solved analytically due to the powerful advantages of the large-group monopolistic-competition assumption. Equations (12.7) and (12.8) can be solved for both X and p_x. Then these solution values can be used in (12.11) to get n. The solution values are:

$$X = (\sigma - 1)\frac{fc_x}{mc_x} \qquad n = \frac{\beta L}{\sigma fc_x} \qquad nX = \frac{(\sigma - 1)}{\sigma}\frac{\beta L}{mc_x} \tag{12.13}$$

Note from the first equation of (12.13) that the output of any good that is produced is a constant and that from the second equation that any expansion in the economy creates a proportional increases in variety n. Let X/L, the consumption of a representative variety per capita, be given by C. Then note from the last equation of (12.13) that nC is a constant:

$$nC = n\frac{X}{L} = \frac{(\sigma - 1)}{\sigma}\frac{\beta}{mc_x} \tag{12.14}$$

Figure 12.3 plots n against C, and equation (12.14) is shown as a negatively-sloped curve in this Figure. Next, note from the second equation in (12.13) that n depends only on L and fixed parameters. Thus we show a second curve which is just a horizontal line in Figure 12.3 which gives the fixed value of n for a given value of L. The intersection of these two curves gives the number of varieties and consumption of a representative variety per capita. For a single economy, the outcome could be shown by variety level n^0 and variety consumption per capita by C^0.

Figure 12.3

Now repeat our usual experiment: put two identical countries together in free trade. This is represented by simply letting L double. The nC curve in Figure 12.3 does not shift as shown by (12.14), but the n curve shift up in proportion to L, doubling in value. If n doubles, then from (12.14) C must be cut in half. The new values in the open economy are C^1 and n^1 in Figure 12.3. Note that this is exactly what is analyzed in (12.3) and suggested in Figure 12.2 above. Equations (12.7) and (12.10) will show that consumption per capita of Y remains unchanged after trade, and hence welfare increases in each country due to the variety effect.

12.3 Monopolistic competition in specialized intermediate inputs (basics idea, then algebra after (12.20) can be skipped without loss of continuity)

The basic idea behind love of variety has also been applied to intermediate inputs, starting with Ethier (1982). This has in turn been applied in a number of different context including endogenous growth models, technology transfer through trade, and trade in producer services. Here we will analyze a much simplified Ethier model, in particular as extended to consider trade in final goods only versus trade in intermediates in Markusen (1989).

Suppose that there are two final consumption goods, X and Y, which are homogeneous and produced with constant returns to scale by competitive firms. Utility of the representative consumer is given by

$$U = U(X, Y) \tag{12.15}$$

There is a factor of production labor, L, which is in fixed supply. In addition, we assume a sector-specific factor K in the Y sector. The purpose of K is to generate a concave (bowed out) production fontier as we will discuss shortly. Good X is assumed to be costlessly assembled from differentiated or specialized intermediate goods S_i in a Dixit-Stiglitz fashion. The two production functions are given as follows, where σ is the elasticity of substitution as derived earlier in the book.

$$Y = G(L_y, \bar{K}) \qquad X = \left[\sum_i^n S_i^\beta\right]^{1/\beta} \qquad \sigma = \frac{1}{1-\beta} \tag{12.16}$$

Each S_i is produced with increasing returns to scale, consisting of the constant marginal cost and fixed-cost technology that we have now used many times. To reduce notation, one unit of S requires a single unit of labor. Labor requirements in S goods and the total labor supply constraint are given as follows, where n is the (endogenous) number of intermediates.

$$L_{xi} = wS_i + wF \qquad \bar{L} = L_y + nL_{xi} \tag{12.17}$$

Since each S enters (12.16) symmetrically and each has an identical technology, we can anticipate the result that any S that is produced is produced in the same amount and sells for the same price as any other S. Let superscript "a" denote a situation in which only the final X and Y goods can be traded and n^a the number of and S^a the amount of each S good in the "a" equilibrium. The X technology reduces to

$$X^a = \left[\sum_i^{n^a} (S_i^a)^\beta\right]^{1/\beta} = \left(n^a\right)^{1/\beta} S^a \tag{12.18}$$

Now again do our standard experiment where we put two identical economies together in trade. In order to illustrate the main idea, hold the amount of each S good produced constant and assume the number produced in each country constant. The number of intermediate goods used in each country in free trade in intermediates is double the number in autarky *(n^* = na)*, with the total output of each shared evenly between the two countries. Output of X in each country is now given by

$$X^* = \left[\sum_i^{n^*} (S_i^a/2)^\beta\right]^{1/\beta} = \left(2n^a\right)^{1/\beta}(S_i^a/2) \tag{12.19}$$

Simplifying the right-hand side, we can compare X output with intermediates trade to output under trade in final goods only given in (12.18).

$$X^* \;=\; 2^{\frac{1-\beta}{\beta}} (n^a)^{1/\beta} S_i^a \;=\; 2^{\frac{1-\beta}{\beta}} X^a \;>\; X^a \tag{12.20}$$

Allowing trade in intermediates increases productivity in *X* production as *X* producers now have access to a larger range of specialized intermediates, a greater division of labor.

Results are shown in Figure 12.4, where we assume that the diminishing returns to the fixed factor *K* in the *Y* sector outweigh the increasing returns to scale in the *X* sector, so that the production frontier is concave. The frontier through A gives the frontier when only final goods can be traded. Both countries are identical by assumption, and so there are no gains from trade: countries could trade but there is no benefit from doing so. Point A could represent the equilibrium production and consumption point for each country under goods trade only (the non-tangency of the price ratio with the frontier will be treated shortly).

Figure 12.4

If we do allow trade in intermediate goods (which could include specialized services), then the production frontier shifts to the one passing though F in Figure 12.4. Each country exports half of each of its inputs for half of each of the other country's inputs. With free trade in intermediatates, production and consumption for each country could be at a point like F in Figure 12.4.

As noted earlier, this formulation of the monopolistic competition model in a manner reminiscent of Adam Smith's division of labor has been used in a number of contexts including endogneous growth theory and the liberalization of trade in producer services. We now look at the issue of optimality of the market outcomes in this model; this is a somewhat more esoteric issue and the remainder of the section may be skipped by some readers. Results of this are equally applicable to the more standard final-goods model treated in the previous two sections.

Suppose that the economy faces a fixed price of *X* relative to *Y*, denoted *p*. Then the optimal number of intermediates and the output level of each can be found by maximizing the value of final output of *X* minus input costs (representing the opportunity cost of labor in producing *Y*.

$$Max\ \pi^* \;=\; p\left[\sum S_i^{\beta}\right]^{1/B} - \sum\,(wS_i + wF) \quad \text{with respect to } n \text{ and } S_i\text{, for all i} \tag{12.21}$$

The first-order condition with respect to S_i is given by applying the chain rule.

$$\frac{\partial \pi^*}{\partial S_i} \;=\; (p/\beta)\left[nS_i^{\beta}\right]^{\alpha}\beta S_i^{\beta-1} - w \;=\; pn^{\alpha} - w \;=\; 0 \qquad \alpha \;\equiv\; (1-\beta)/\beta \tag{12.22}$$

The first-order condition for the number of goods is given by the effect of adding one more *n*.

$$\frac{\partial \pi^*}{\partial n} \;=\; (p/\beta)\left[nS_i^{\beta}\right]^{\alpha} S_i^{\beta} - wS_i - wF \;=\; (p/\beta)n^{\alpha}S_i - wS_i - wF \;=\; 0 \tag{12.23}$$

If we solve these two equations in two unknowns, we get the optimal output of any intermediate that is produced. Note that w is just the marginal product of labor in the *Y* sector, G_l. From (12.22), we can also get the optimality condition for a tangency between the price ratio and the marginal rate of transformation along the production frontier.

$$S_i = \left(\frac{\beta}{1-\beta}\right)F \qquad p = \frac{w}{n^{\alpha}} = \frac{G_l}{n^{\alpha}} = MRT \tag{12.24}$$

Now turn to the outcome under a market solution. The price of an individual *S* is the value of its marginal product in producing *X*. This is in fact given in the first term on the right-hand side of the first equation in (12.22). Let *r* denote the price of an individual *S*.

$$r = (p/\beta)\left[nS_i^{\beta}\right]^{\alpha}\beta S_i^{\beta-1} = qS_i^{\beta-1} \qquad q \equiv p\left[nS_i^{\beta}\right]^{\alpha} \tag{12.25}$$

In the tradition of large-group monopolistic competition discussed above, assume that each individual *S* producer views q in (12.25) as fixed. Then each *S* producer maximizes the following expression with resepect to S_i viewing *q* as fixed.

$$Max\ \pi = (qS_i^{\beta-1})S_i - wS_i - wF \tag{12.26}$$

The first-order condition is given by

$$\frac{\partial\pi}{\partial S_i} = q\beta S_i^{\beta-1} - w = 0 \tag{12.27}$$

Secondly, the free entry condition to determine *n* is that each *S* producer's profits are zero.

$$qS_i^{\beta} - wS_i - wF = 0 \tag{12.28}$$

Solving (12.27) and (12.28), we get the market equilibrium amount of any *S* that is produced. Second, substituting the expression for *q* in (12.25) back into (12.27), we can get the relationship between the competitive price ratio and the marginal rate of transformation given in (12.24). These are given by

$$S_i = \left(\frac{\beta}{1-\beta}\right)F \qquad p\beta = \frac{w}{n^{\alpha}} = MRT < p \tag{12.29}$$

Comparing the optimum in (12.24) to the market outcome in (12.29) we see that any *S* that is produced is produced in the optimal amount. But we also see that the market outcome involves a distortion between the price ratio and the marginal rate of transformation, much as in the case of a tax, or the external-economies case of Chapter 10, or the oligopoly outcomes in Chapter 12. This is the distortion between the price ratio and the slope of the frontier shown in Figure 12.5.

The conclusion of this exercise is that the market outcome is not first best: it produces the optimal output of any good that is produced, but too few goods are produced. The intuition behind this is essentially an externality argument. (12.28) gives the private profits of entry. But when one firm enters, it increases the productivity of every other firm holding prices constant. This is seen in (12.25): the value of the marginal product of an additional unit of *S* in producing *X* (*r*) is increasing in n, the division of labor. This effect is not considered by an individual firm in its entry decision and hence there is a positive externality in the *X* sector. Note that when (12.28) holds with equality (the private profits from entry are zero), the "social" marginal product of an addition *S* given in (12.23) is strictly positive. There is thus a close analogy here between the present result and that in Chapter 10 in the section on production externalities.

12.4 The ideal variety approach to product diversity

Gains from trade through product differentiation can be looked at in a second way as well. While consumers may prefer diversity as just noted, consumers themselves may also have different tastes. Consumers may, for example, buy only one automobile each, but they have different views as to what is the "ideal" automobile for their tastes and income level. This approach to product diversity is thus labelled the "ideal variety" approach (Lancaster, 1980). Due to scale economies, no country can afford to produce a unique automobile for each consumer. Germany produces Volkswagens and Mercedes, and France produces Renaults and Peugeots, all of which have somewhat different characteristics from consumers' points of view. Trade in automobiles then occurs between France and Germany due to the fact that some Germans prefer Renaults or Peugeots and some Frenchmen prefer Volkswagens or Mercedes.

This situation is shown in Figure 12.5. Suppose that automobiles have only two characteristics: C_1 and C_2 (e.g., size and speed). There is a trade-off between these two characteristics such that if one wants a bigger car he or she must sacrifice some speed if the two are going to cost the same (e.g., Mercedes versus Porsche). Figure 12.5 shows three possible combinations of C_1 and C_2, denoted X_1, X_2, and X_3, corresponding to three different types of cars. Suppose that all three models could be produced at the same average cost for the same volume of production and that at this common cost, the amounts of C_1 and C_2 embodied in the cars are given by points (X_2^0, X_3^0, X_1^0) in Figure 12.5. The straight line through these points represents a sort of iso-cost line at a common scale of production.

Figure 12.5

Now suppose that we have two groups of consumers with distinct tastes (identical within each group). Consumer type-1 has a relative preference for characteristic C_1 and consumer type-2 has a relative preference for characteristic C_2. On the isocost line shown in Figure 12.5, an indifference curve for consumer type-2 is tangent to the isocost line at point X_2^0 and an indfference curve for consumer type-1 is tangent at point X_1^0. X_2 is then referred to as consumer type-1's *ideal variety* and X_1 as consumer type-1's ideal variety.

Variety X_3 could be referred to as a *compromise variety*. Note from Figure 12.5 that for X_3 to be equally attractive to our two consumer types, more "stuff" (e.g., stereo, air conditioning) would have to be added to it: to make the consumers indifferent between the compromise variety and their ideal varieties, we would have to offer X_3^1, which lies outside the isocost line. The proportional difference $(X_3^1 - X_3^0)/X_3^0$ in Figure 12.5 is sometimes called the *compensation ratio*: the proportional amount of extra "stuff" needed to make the compromise variety as attractive as the ideal varieties.

If there were constant returns to scale, the problem would be trivial: each consumer would get there ideal variety, no one would even buy the compromise variety because it would cost more. However, the problem is far from trivial with increasing returns to scale. The production scale in producing the compromise variety is twice as large as in giving each consumer their ideal variety. Thus, adjusting for the benefits of larger production scale, variety , X_3^1, may in fact cost less than variety X_2^0 or X_1^0.

Suppose that there is only a single factor labor, L, with a wage one. We assume, as is typical in this approach, that each consumer wants only one unit of the good but of course prefers their ideal variety. There are fixed and marginal costs for each variety. mc0 is the labor needed for on unit of (X_2^0, X_3^0, X_1^0) and mc1 is the labor need for one unit of X_3^1. Labor requirements are then

$$L_1^0 = L_2^0 = L_3^0 = mc^0 + fc \qquad L_3^1 = mc^1 + fc \qquad mc^0 < mc^1 \tag{12.30}$$

The relevant question is whether or not two ideal varieties (two times the first equation of (12.29)) costs more than two units of the compromise variety. The compromise variety is preferred if

$$2mc_3^1 + fc < 2mc_i^0 + 2fc \quad or \quad 2(mc_3^1 - mc_3^0) < fc \quad i=1,2,3 \tag{12.31}$$

The important point is that the compromise variety means incurring the fixed costs only once versus twice if each group gets its ideal variety.

Suppose that the inequality in (12.30) "marginally" holds, meaning that the left-hand side is less than the right-hand side by a very small amount. Then the compromise variety is preferred. Consider then our now-familiar experiment of putting two identical countries together. To give every consumer one unit, the total world labor requirements for ideal varieties are now twice the marginal costs shown in (12.30) and similarly for the compromise variety: no additional resources are needed for fixed costs. Total requirements (costs) and the condition for the compromise variety to be preferred are now

$$4mc_3^1 + fc < 4mc_i^0 + 2fc \quad or \quad 4(mc_3^1 - mc_3^0) < fc \quad i=1,2,3 \tag{12.32}$$

If (12.30) holds marginally, the inequality in (12.31) will *not* hold. It will be reversed and the ideal varieties will be preferred.

As in the love-of-variety approach, we see here that the ideal-variety approach means that there are gains from trade through increased product variety. The nature of those gains are different here. The idea is that trade allows each consumer type to get a product closer to their ideal variety embodying their ideal characteristic combination. In more general models with many consumer types and goods, the welfare statement in these models is typically that "on average" consumers get varieties closer to their ideal type.

In a world of imperfect competition that accompanies scale economies, we have to be cautious and note that just because something is preferred doesn't mean it is going to happen. These ideal-variety models get quite mathematically complicated, much more so than the simple love-of-variety model. For example, there may be a continuum of consumers who preferences are "located" at difference points of a line or a circle. Such models are sometimes referred to as "location" or "address" models of product differentiation, but these embody the same basic ideas: a consumer's location or address on the circle is the consumer's ideal variety (see Helpman, 1981). These models have been perhaps more used in industrial-organization economics. The problem for international trade is that they quickly get very difficult in general-equilibrium settings.

12.5 Some useful algebra for Dixit-Stiglitz (may be skipped without loss of continuity but useful later for analyzing trade costs)

This section lays out the algebra required to solve for the demand function and price index (defined and discussed shortly) for the Dixit-Stiglitz utility function. In order to simplify it a bit, let's assume that there is just a single section producing differentiated X varieties. But there is a straightforward extension to a two-sector model in which the other sector, Y, has constant returns to scale and perfect competition, and there is Cobb-Douglas substitution between the X goods and good Y. In this situation, the consumer always devotes constant share of income to each sector, so we will start with an allocation of income M_x to the X sector.

Let X_c denote the utility derived from the X varieties; X_c is sometimes referred to as a composite commodity. X_c does have a price associated with it. This is a price index, the minimum expenditure necessary to buy one unit of composite good X_c. We will denote this price or rather price index as e_x. The value of X_c (the utility from X consumption) is defined as follows.

$$X_c = \left[\sum_i^n X_i^\beta\right]^{\frac{1}{\beta}} \qquad \sigma = \frac{1}{1-\beta}, \quad \beta = \frac{\sigma-1}{\sigma}$$

The consumer maximizes utility subject to a budget constraint. The maximization problem is written as a Lagrangean function and the first-order condition for an arbitrary good X_i are given as follows.

$$\max X_c = \left[\sum X_i^\beta\right]^{\frac{1}{\beta}} + \lambda(M_x - \sum p_i X_i) \quad => \quad \frac{1}{\beta}\left[\sum X_i^\beta\right]^{\frac{1}{\beta}-1}\beta X_i^{\beta-1} - \lambda p_i = 0 \qquad (12.33)$$

Divide the first-order condition in (12.33) for an arbitrary good i by the corresponding condition for good j.

$$\left[\frac{X_i}{X_j}\right]^{\beta-1} = \frac{p_i}{p_j} \qquad \frac{X_i}{X_j} = \left[\frac{p_i}{p_j}\right]^{\frac{1}{\beta-1}} = \left[\frac{p_i}{p_j}\right]^{-\sigma} \qquad since \quad \sigma = \frac{1}{1-\beta} \qquad (12.34)$$

Now perform several steps. (1) write the second equation of (12.34) with X_j on the left. (2) multiply both sides of this equation by p_j. (3) sum this equation over all goods j. These three steps are

$$X_j = \left[\frac{p_i}{p_j}\right]^{\sigma} X_i \qquad p_j X_j = p_j p_j^{-\sigma} p_i^{\sigma} X_i \qquad \sum p_j X_j = M_x = \left[\sum p_j^{1-\sigma}\right] p_i^{\sigma} X_i \qquad (12.35)$$

Inverting this last equation, we have the demand for an individual variety i:

$$X_i = p_i^{-\sigma}\left[\sum p_j^{1-\sigma}\right]^{-1} M_x \qquad \sigma = \frac{1}{1-\beta}, \quad \beta = \frac{\sigma-1}{\sigma} \qquad (12.36)$$

Now we can use X_i to construct X_c and then solve for e_x, the price index, noting the relationship between α and σ. First, raise (12.36) to the power β.

$$X_i^\beta = X_i^{\frac{\sigma-1}{\sigma}} = p_i^{1-\sigma}\left[\sum p_j^{1-\sigma}\right]^{\frac{1-\sigma}{\sigma}} M_x^\beta \qquad (12.37)$$

Now sum over all of the i varieties of X.

$$\sum X_i^\beta = \left[\sum p_i^{1-\sigma}\right]\left[\sum p_j^{1-\sigma}\right]^{\frac{1-\sigma}{\sigma}} M_x^\beta = \left[\sum p_j^{1-\sigma}\right]^{\frac{1}{\sigma}} M_x^\beta \qquad (12.38)$$

Now raise both sides of this equation to the power 1/β to get the composite commodity demand for X_c.

$$X_c = \left[\sum X_i^{\beta}\right]^{\frac{1}{\beta}} = \left[\sum X_i^{\beta}\right]^{\frac{\sigma}{\sigma - 1}} = \left[\sum p_j^{\,1-\sigma}\right]^{\frac{1}{\sigma - 1}} M_x = M_x / e_x \tag{12.39}$$

The demand for X_c must be the expenditure M_x on X_c, divided by the price index e_x. Examining the last equation of (12.39), this means that the price index is given by:

$$e_x = \left[\sum p_j^{\,1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{12.40}$$

Again, the price index e_x, also called the expenditure function, is the minimum cost or expenditure necessary to buy one unit of the composite commodity X_c. If all of the X varieties sold for the same price, the expenditure function (price index) in (12.40) simplifies as follows.

$$e_x = n^{\frac{1}{1-\sigma}} p \tag{12.41}$$

Note that the price index is homogeneous of degree one in the prices of the individual goods: if we double all prices we double to cost of buying one unit of X_c. However note that the price index is decreasing in the number of varieties available (1 - σ < 0). The is another way of thinking about the love-of-variety effect. The same utility derived from two apples or two oranges might be derived from (for example) 0.8 apples and 0.9 oranges. When more variety is available, the consumer can achieve the same utility (same value of X_c) by actually reducing total expenditure.

Finally, having derived e, we can then use equation (13) in (9) to get the demand for an individual variety.

$$X_i \equiv p_i^{-\sigma} e_x^{\sigma - 1} M_x \qquad \textit{since} \qquad e_x^{\sigma - 1} = \left[\sum p_j^{\,1-\sigma}\right]^{-1} \tag{12.42}$$

We now move onto a chapter introducing the existence of trade costs. The price index in (12.40) and the demand function in (12.42) will prove very useful in discussing trade costs in the context of monopolistic-competition models in the next chapter.

12.6 Summary

This is the second of two chapters on trade with increasing returns to scale and imperfect competition. The first (Chapter 11) focused on a pure case in which firms produced identical goods but scale economies limited the number of firms such that individual firms took into account their strategic interactions with other firms in their output decisions. A principal result is that the larger economy created through trade offer welfare gains through pro-competitive and production-scale effects. This Chapter focused also on a pure case in which firms produce somewhat different goods and there are a large number of firms such that each views the decision of rival firms as exogenous. A principal result is that the larger economy created through trade offers welfare gains through increased product or intermediate-good diversity: more products at the same prices rather than the same products at lower prices in Chapter 11. Of course, the two approaches can be combined, but the analysis becomes more difficult and is left to more advanced treatments (e.g., Melitz and Ottaviano (2008)) and indeed the ideal variety approach does involve variable markups and pro-competitive effects.

The two main approaches to product diversity are considered. The “love of variety” approach assume an endogenous set of symmetric but imperfectly substitutable products. Consumers are rewarded by a more diverse consumption bundle through trade. The “ideal variety” approach works rather differently. Consumers are assumed to differ in their views as to the ideal product, a product being a bundle of characteristics. Often in this approach, the consumer is assumed to buy just a single unit or nothing. As in the love-of-variety approach, diversity, in this case giving each consumer their ideal product, is costly in the presence of scale economies and so consumers accept compromise varieties in small autarky markets. The rewards to trade are that consumers, on average, get products closer to their ideal varieties.

While much of the literature has focused on final goods, an important variation of these models considers the differentiated products to be intermediate goods used in producing homogeneous final goods. Allowing trade in intermediate goods offers final producers higher productivity by increasing the division of labor. This ideal is a cornerstone of what is referred to as endogenous growth theory (e.g., Romer 1987) and has also been applied to trade in components and in producer services.

REFERENCES

Dixit, Avinash K. and Joseph Stiglitz (1977), “Monopolistic Competition and Optimum Product Diversity”, *American Economic Review* 67, 297-308.

Ethier, Wifred J., (1982), "National and International Returns to Scale in the Modern Theory of International Trade", *American Economic Review* 72, 389-406.

Helpman, Elhanan, (1981). "International Trade in the Presence of Product Differentiation, Economies of Scale and Monopolistic Competition. A Chamberlinian, Heckscher-Ohlin Approach." *Journal of International Economics* 11, 304-340.

Helpman, E. and P. Krugman (1985). *Market Structure and Foreign Trade*, Cambridge: MIT press.

Krugman, P. (1979). "Increasing Returns, Monopolistic Competition, and International Trade." *Journal of International Economics* 9, 469-479.

Krugman, P. (1981), "Intraindustry Specialization and the Gains from Trade", *Journal of Political Economy* 89, 469-479.

Lancaster, K. (1980). "Intra-Industry Trade Under Perfect Monopolistic Competition." *Journal of International Economics* 10, 151-175.

Markusen, J. R. (1989). Trade in Producer Services and in Other Specialized Intermediate Inputs." *American Economic Review* 79, 85-95.

Melitz, Mark J. and Gianmarco I.P. Ottaviano (2008), “Market size, trade, and productivity”, *Review of Economic Studies* 75, 295-298.

Romer, Paul M. (1987), “Growth Based on Increasing Returns to Scale due to specialization”, *American Economic Review* 77, 56-62.

Figure 12.1

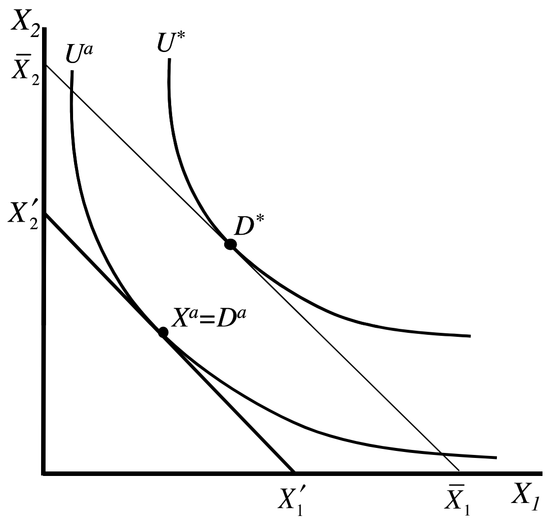


Figure 12.2

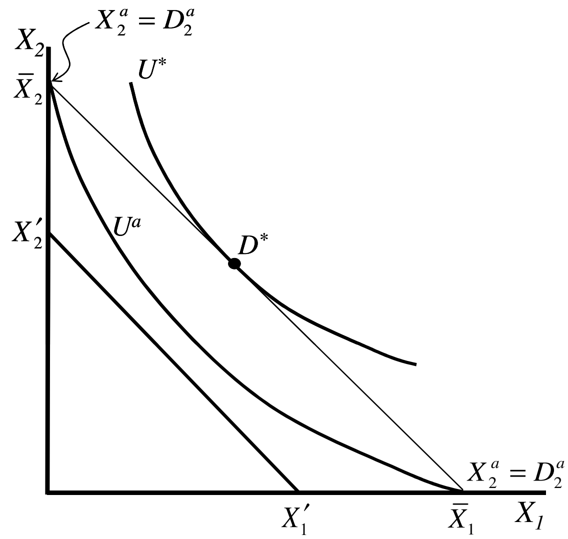


Figure 12.3

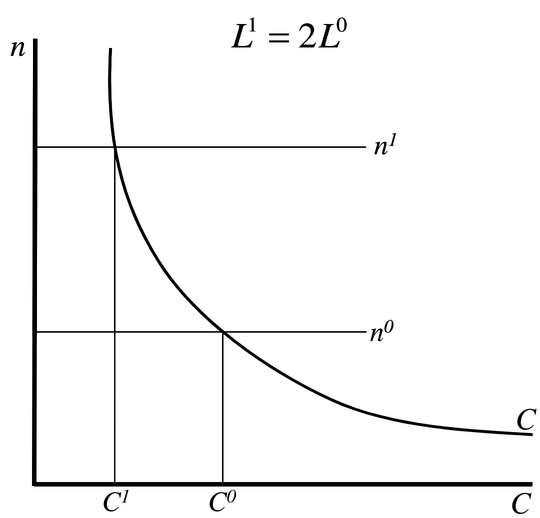


Figure 12.4

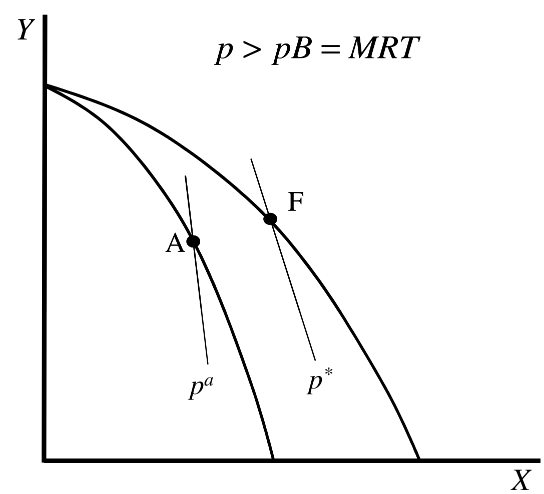
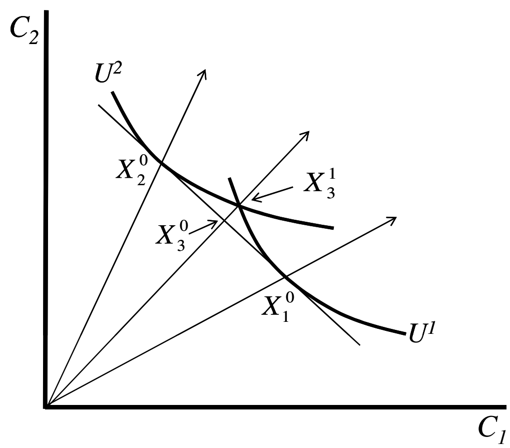


Figure 12.5



Chapter 16

MULTINATIONAL FIRMS AND FOREIGN DIRECT INVESTMENT

16.1 Stylized facts, basic concepts

Multinational firms have become a crucial element in the modern world economy, and a few statistics are presented in Table 16.1 to document this. The first number documents that fact that sales of foreign affiliates of multinational corporations are double the total value of world trade. Other number we have seen range up to a factor of five. The second number notes that value added in multinational affiliates is about 11 percent of world GDP (itself a value-added measure). The third number notes that one-third of all world exports originate in the foreign affiliates of multinational firms. The final number actually surprises some folks as being rather small: foreign affiliate exports are only 18 percent of their total sales. Most of the output of foreign affiliates is sold locally, a point that we very much need to make sense of.

Table 16.1

Several decades ago, foreign direct investment (FDI) was viewed as simply a capital movement. Consistent with the dominant theory of the day, Heckscher-Ohlin theory, capital should flow from capital-rich high-income countries to poor capital-scarce countries. During the 1980s, it became very obvious that the old view was at best inadequate and at worst simply wrong

(A) FDI flows primarily from high-income developed countries to other high-income countries, not from capital-rich to capital-poor countries.

Table 6.2 gives some evidence on this. The top panel gives flows of new FDI for 2007, expressed as the *developed* countries' share of total flows, in both inward and outward directions. These figures are followed by the corresponding shares for the stock of FDI in 1990 and 2007. Clearly, the developed countries are the major source of FDI, but what is less appreciated is that they are also the major recipients of FDI.

Table 16.2

(B) Affiliate production is primarily for local sale and not for export back to the parent country.

Another myth that persists in the popular press though I hope not in the economics literature is that multinationals move production to poor countries to pay low wages and export the output back home. Table 6.3 breaks down sales of foreign manufacturing affiliates of US, Japanese, and Swedish parents into local sale and export sales, and imports are also listed. Note that two-thirds or more of foreign output of affiliates is sold locally. Table 16.4 has data from a more restricted sample for the US and Sweden. In 2003, local sales account for 60 percent of all sales, but that still leaves a large portion for export. But columns 2 and 3 reveal that most of the exports do not go back to the parent country, they go to third countries. The point is that foreign affiliates are not primarily in the business of producing cheaply abroad for shipment back to the parent: they are primary in the business of producing abroad for local and regional markets.

Table 16.3 **Table 16.4**

Several other stylized facts will be important.

(C) FDI is attracted to large markets and high-income markets

This again fits well with the idea that production is largely for local or regional sale and, if there are fixed costs of setting up or acquiring new plants, then such investments are more likely to be observed in large and high-income markets. If multinationals were simply in search of low wages to produce for export back home, then we should not observe firms to choose high-income markets and there should be little relationship to host-country size.

(D) There are high levels of intra-industry cross-investment, particularly among the high-income countries

Much has been written about intra-industry trade or "cross-hauling" over the last several decades. It seems to be less well known that exactly the same phenomenon is observed for affiliate sales. Combined with earlier statistics, these numbers emphasize that a satisfactory theory of multinational firms must be consistent with large volumes of cross-investment (intra-industry affiliate production) among similar large, high-income countries. But it must also of course, be able to explain the fact that firms from the high-income countries are net investors in developing countries.

The weight of these statistics also suggests that much if not most FDI is "horizontal" or "market seeking"; that is, foreign affiliates of multinational firms are doing much the same things in foreign countries as they are at home. The same products and services are produced, generally for local and regional sale. "Vertical" or "resource seeking" (e.g., low cost labor) investments involve the fragmentation of the production process into stages, with stages located where the factors of production they use intensively are relatively cheap. While not the dominant motivation for FDI, vertical investments may nevertheless be quite important in developing countries.

16.2 A basic organizing framework

Modern theory often begins with the premise that firms incur significant costs of doing business abroad relative to domestic firms. Therefore, for a firm to become a multinational, it must have offsetting advantages. A limited but very useful organizing framework for inquiring into the nature of these advantages was proposed by John Dunning (1977). Dunning proposed that there are three conditions needed for firms to have a strong incentive to undertake direct foreign investments.

Ownership Advantage: the firm must have a product or a production process such that the firm enjoys some market power advantage in foreign markets.

Location Advantage: the firm must have a reason to want to locate production abroad rather than concentrate it in the home country, especially if there are scale economies at the plant level.

Internalization Advantage: the firm must have a reason to want to exploit its ownership advantage internally, rather than license or sell its product/process to a foreign firm.

Internalization is often nowadays referred to as vertical integration and the choice between and owned subsidiary versus licensing is often now referred to as the outsourcing decision, which is the same thing put the other way around.

The basic idea is shown in Figure 6.1. A firm wanting to produce in a foreign country for local sale has a location decision between producing there and exporting. We can refer to this as the "offshoring" decision. If the multinational chooses foreign production, it has the further choice between an owned subsidiary, perhaps some sort of joint venture, and an arm's length licensing contract. We can refer to this as the internalization or vertical integration versus outsourcing decision, whether or not to produce outside the ownership boundaries of the firm.

Figure 16.1

Consider first ownership advantages. Evidence indicates that multinationals are related to R&D, marketing, scientific and technical workers, product newness and complexity, and product differentiation (Caves 2007, Markusen 2002). This suggests that multinationals are firms which are intensive in the use of knowledge capital. This is a broad term which includes the human capital of the employees, patents, blueprints, procedures, and other proprietary knowledge, and finally marketing assets such as trademarks, reputations, and brand names.

There are several reasons why the association of multinationals with knowledge-based assets rather than physical capital is appealing. First, the services of these assets may be easily used in distant plants, such as managers and engineers visiting those plants. Second and more subtlety, knowledge capital often has a joint-input or non-rivaled property within the firm. Blueprints, chemical formulae, or even reputation capital may be very costly to produce, but once they are created, they can be supplied at relatively low cost to foreign production facilities without reducing the value or productivity of those assets in existing facilities.

The sources of location advantages are varied, primarily because they can differ between horizontal and vertical firms. Consider horizontal firms that produce the same goods and services in each of several locations. Given the existence of plant-level scale economies, there are two principal sources of location advantages in a particular market. The first is the existence of trade costs between that market and the MNEs home country, in the form of transport costs (both distance and time), tariffs and quotas, and more intangible proximity advantages. The second source of location advantage, again following from the existence of plant-level scale economies, is a large market in the potential host country. If that market is very small, it will not pay for a firm to establish a local production facility and the firm will instead service that market by exports.

The sources of location advantage for vertical multinationals are somewhat different. This type of investment is likely to be encouraged by low trade costs rather than by high trade costs and by factor-price differences across countries. Low trade costs facilitate the intra-firm trade of intermediate and final goods and thus facilitate the geographic breaking up of the production chain.

Internalization advantages (or outsourcing disadvantages) are the most abstract of the three. The topic quickly gets into fundamental issues such as what is a firm, and why and how agency problems might be better solved within a firm rather than through an arm's-length arrangement with a licensee or contractor. Basically, it is our view that internalization advantages often arise from the same joint-input, non-rivaled property of knowledge that create ownership advantages.

16.3 A simple monopoly model of location choice

In this section, we explore a basic model very similar to that used in Chapters 11 and 13 in order to get the crucial intuition as to the optimal international organization of a firm. Questions of outsourcing versus vertical integration (internalization) are left to a later section. There are two countries, home and foreign and one monopoly firm in country h. There is a linear inverse demand for the product where the intercept is α and slope is (*1/L*), L = market size. The price (p_i), quantity (X_i) and market size (L_i) in market i = h,f are related as follows, where the second equation is firm revenues (R_i) in market i.

$$p_i = \alpha - X_i/L_i \qquad R_i = p_i X_i = (\alpha - X_i/L_i)X_i \tag{16.1}$$

There is a constant marginal cost c_i in market i and a specific trade cost t between markets. Profits before fixed costs for a plant producing in market i and selling in i and a plant producing in i and selling in j are given by

$$\pi_{ii} = (\alpha - X_{ii}/L_i)X_{ii} - c_i X_{ii} \qquad \pi_{ij} = (\alpha - X_{ij}/L_j)X_{ij} - (c_i + t)X_{ij} \tag{16.2}$$

Taking the first-order conditions for profit maximization given the optimal levels of domestic and export supply.

$$X_{ii} = \left(\frac{\alpha - c_i}{2}\right)L_i \qquad X_{ij} = \left(\frac{\alpha - c_i - t}{2}\right)L_j \tag{16.3}$$

The firm is headquartered in country h but may chooses between three (discrete) alternatives. It can choose a single plant at home in country h, exporting to country f and we refer to this as a national or domestic firm (d). It can have a single plant in country f, exporting back to h, a vertical (v) structure. Or it can be a horizontal multinational with plants in both countries (m). There is a firm-specific fixed cost F and plant specific fixed cost G where the latter must be incurred for each plant. This model thus displays firm-level scale economies: the total fixed costs of a two-plant firm are *(F + 2G)* which is less that the total fixed costs of two independent single-plant firms, *(2F + 2G)*. F thus represents the cost of creating a product or process, and this knowledge is a joint or non-rivaled input across plants.

If you substitute (16.3) into the two equations of (6.2) and add in fixed costs, you will find that the profits for each of the multinational firms three choice are given as follows (we did a derivation of this in Chapter 13).

Profits of a national firm: one plant at home (h) exporting to f

$$\Pi^{\ddot{\upsilon}} = \Pi_{\vartheta\vartheta} + \Pi_{\vartheta\acute{\omega}} = \left[\frac{\alpha - c_{\vartheta}}{2}\right]^{\theta} L_{\vartheta} + \left[\frac{\alpha - c_{\vartheta} - t}{2}\right]^{\theta} L_{\acute{\omega}} - G - F \tag{16.4}$$

Profits of a vertical firm: one plant in j exporting back to i

$$\Pi^{\prime} = \Pi_{\acute{\omega}\vartheta} + \Pi_{\acute{\omega}\acute{\omega}} = \left[\frac{\alpha - c_{\acute{\omega}} - t}{2}\right]^{\theta} L_{\vartheta} + \left[\frac{\alpha - c_{\acute{\omega}}}{2}\right]^{\theta} L_{\acute{\omega}} - G - F \qquad (16.5)$$

Profits of a horizontal firm: plants in both countries.

$$\Pi^{\Upsilon} = \Pi_{\vartheta\vartheta} + \Pi_{\acute{\omega}\acute{\omega}} = \left[\frac{\alpha - c_{\vartheta}}{2}\right]^{\theta} L_{\vartheta} + \left[\frac{\alpha - c_{\acute{\omega}}}{2}\right]^{\theta} L_{\acute{\omega}} - 2G - F \qquad (16.6)$$

This is really a surprisingly rich little model. Figure 16.2 shows profits from the three possible choices holding total world demand constant, but varying the size of the two markets along the horizontal axis with the parent country h small on the left and big on the right. If you stare at these three equations long enough and consider how each curve shifts (or does not shift) in response to some parameter change, you will conclude that a two-plant *horizontal* structure is more likely as:

Both markets are large	characteristic of markets
Markets of similar size	characteristic of markets
Marginal costs are similar	characteristic of markets
Firm fixed costs > plant fixed costs	characteristic of industry
Transport/tariff costs are large	geography/policy

Figure 16.2

Large markets mean that the added fixed costs of a second plant outweigh the higher variable costs of exporting. The intuition behind the second and third results is that if one market is much larger and/or production costs much smaller, then it pays to put a single plant there and export to the smaller/costlier market. For the third result, note that if we raise *F* and lower *G* in the same amounts, this increases the profits of a type-m firm while leaving the profits of a type-d or type-v firm unchanged (the horizontal curve shifts up). Trade costs reduce the profits of a type-d or type-v firm but leave the profits of a type-m firm unchanged (the national and vertical curves shift down).

Despite its simplicity, this model fits the data well: horizontal firms will be important between similar, large (rich) markets in industries where knowledge capital is important (*F* is large relative to *G*).

A vertical structure is preferred to a national structure as:
Foreign market is larger
Foreign marginal cost is low
Low trade costs: vertical structure if $c_f < c_h$ even if country f is very small

As noted above, if you are going to have a single plant, put it in the large and/or low-cost market. Note however that the importance of the local market size disappears as trade costs go to zero. As t converges to zero, the relationship between (16.4) an (16.5) is determined entirely by which is the low-cost location, and note that the horizontal structure will never be chosen at $t = 0$ ($G > 0$). As t goes to zero, we can say that the vertical structure is chosen if and only if country f is the low-cost location.

16.4 Monopolistic competition and the choice of exporting versus horizontal production

A crucial feature of the modern theory of the multinational as noted earlier is the notion of joint or non-rivaled knowledge-based assets of firms (these are sometimes also called firm-level or multi-plant economies of scale). Once a firm develops a product, process, or even brand identification and a reputation for quality, it can add additional plants abroad for significantly less additional fixed costs than is incurred for setting up the firm and first plant at home. In this section, we will consider a firm wanting to sell abroad as well as at home, and limit its choices to exporting to the foreign market or becoming a two-plant horizontal multinational. We will use the monopolistic-competition model developed in Chapters 12 and 13.

To keep things simple, we will consider *identical countries*. Assuming in addition that marginal costs of production is the same for both domestic and multinational firms, the pricing equation in the model says that all varieties will have the same (domestic) prices in equilibrium, $p_i = p_j$, regardless of whether they are produced by a locally-owned firm or by a branch plant of a foreign multinational. Similarly, imported goods will have the same prices in each country. Let superscript d denote a domestic (one plant) firm and superscript m denote a two-plant horizontal multinational. As in Chapter 13, $\phi = t^{1-\sigma}$ is the "phi-ness" of trade where t is the (gross) iceberg trade cost: $\phi = 1$ is free trade and $\phi < 1$ means positive trade costs. Let the first subscript on X denote country of firm ownership and the second the country of sale. The demand functions for the various X varieties (domestic production and imports) are given by:

$$X_{ii}^d = X_{jj}^d = X_{ij}^m = X_{ji}^m = p_i^{-\sigma} e_i^{\sigma-1} L/2 \qquad X_{ji}^d = X_{ij}^d = p_i^{-\sigma} \phi e_i^{\sigma-1} L/2 \tag{16.7}$$

Zero-profit conditions for d and m firms located in country i are markup revenues equal fixed costs. Let fc^d denote the fixed costs of a first plant and βfc^d denote the fixed costs of a two-plant multinational. The non-rivaled or multi-plant economies idea implies that $1 < \beta < 2$: a two-plant firm has less than double the fixed costs of a one-plant firm. The zero-profit conditions for one and two-plant firms are then

$$(p_i/\sigma)X_{ii}^d + (p_i/\sigma)X_{ij}^d \le fc^d \tag{16.8}$$

$$(p_i/\sigma)X_{ii}^m + (p_i/\sigma)X_{ij}^m \le fc_x^m = \beta fc^d \tag{16.9}$$

Using the demand functions for X_{ii} and X_{ij} above, these are:

$$p_i^{1-\sigma} e_i^{\sigma-1} L/2 + p_i^{1-\sigma} \phi e_j^{\sigma-1} L/2 \le \sigma fc^d \tag{16.10}$$

$$p_i^{1-\sigma} e_i^{\sigma-1} L/2 + p_i^{1-\sigma} e_j^{\sigma-1} L/2 \le \beta fc^d \tag{16.11}$$

Suppose that we pick values of parameters such that national and multinational firms can both just break even in the two identical countries: (16.10) and (16.11) both hold with equality. Then dividing the first equation by the second give us the critical relationship between trade costs and fixed costs for indifference.

$$\frac{(1+\phi)}{2} = \frac{1}{\beta} \qquad 2 > (1+\phi) = \frac{2}{\beta} > 1 \tag{16.12}$$

Lower trade costs (*higher* ϕ) must be balanced against higher firm-level scale economies (*lower* β) for firms to be indifferent as to type. To put it differently, higher trade costs encourage firms to be multinationals and higher firm-level scale economies do the same. In line with our earlier discussions, we can think of high firm-level scale economies as being associated with knowledge and R&D intensive industries. Our results clearly have empirical predictions about what sort of firms should be multinationals and which industries should be dominated by multinationals. Tests of these ideas have confirmed the basic theory and we will discuss these later.

16.5 The knowledge-capital model

A recent development that integrates both horizontal and vertical motives for multinationals is Markusen's knowledge-capital model. The full definition of the model is found in Markusen (2002). The knowledge-capital model is a general-equilibrium approach that incorporates both horizontal and vertical motives for multinationals. The configuration of firms that arises in equilibrium depends on country characteristics (size, relative size, and relative endowments), industry characteristics (firm versus plant-level fixed costs or scale economies) and trade costs.

There are two goods, *X* and *Y* and two factors of production, skilled and unskilled labor, *S* and *L* There are two countries i and j. *Y* is produced with constant returns by a competitive industry and unskilled-labor intensive. *X* is produced with increasing returns by imperfectly competitive firms. There are both firm-level and plant-level fixed costs and trade costs and firm-level fixed costs result in the creation of "knowledge-based assets".

There are three defining assumptions for the knowledge-capital model.

(A) *Fragmentation*: the location of knowledge-based assets may be fragmented from production. Any incremental cost of supplying services of the asset to a single foreign plant versus the cost to a single domestic plant is small.

(B) *Skilled-labor intensity*: knowledge-based assets are skilled-labor intensive relative to final production.

(C) *Jointness*: the services of knowledge-based assets are (at least partially) joint (non-rivaled) inputs into multiple production facilities. The added cost of a second plant is small compared to the cost of establishing a firm with a single plant.

There are three possible firm "types" that can exist in equilibrium in either country (so six firm types in all), and there is free entry an exist into and out of firm types.

Type m - horizontal multinationals which maintain plants in both countries, headquarters is located in country i or j.

Type d - national firms that maintain a single plant and headquarters in country i or j. Type d_i firms may or may not export to the other country.

Type v - vertical multinationals that maintain a single plant in one country, and headquarters in the other country. Type v_i firms may or may not export back to their headquarters country.

Various assumptions can be make about factor intensities and they do make some quantitative difference to the results. The ones used to generate the diagrams attached below assume that the skilled-labor intensity of activities are

[headquarters only] > [integrated *X*] > [plant only] > [*Y*]

When countries are similar in size and in relative factor endowments, the model predicts that horizontal multinationals will be important. Firms will build plants in the foreign country to serve the local market instead of incurring trade costs on exports. We discussed this in the previous section for identical countries, but now need to comment on this issue of size and endowment similarity. The intuition about the role of these characteristics is understood by considering what happens when countries are *dissimilar* in size or in relative endowments.

To a good degree, the intuition is seen in Figure 16.2 above. When one country is quite large and the other quite small, a national firm locating in the large country will have an advantage since is spends little in trade costs whereas a horizontal firm still has to spend a lot on a second plant regardless of the size of the market. Similarly, if the countries are similar in size but one country is skilled-labor abundant, a vertical firm locating in that country has an advantage over the other two types: it can locate the headquarters in the skilled-labor-abundant country and the single plant in the unskilled-labor-abundant country. Note that this advantage for the vertical firm is reinforced if the skilled-labor-abundant country is also small. The vertical firm has the added advantage that locating its single plant in the large market abroad economizes on trade costs.

Figure 16.3 plots the results of a simulation solving for the number, types, and production of firms over a world Edgeworth box. The world skilled-labor endowment is on the axis running toward the northwest and the unskilled-labor endowment on the axis running to the northeast. The total equilibrium volume of affiliate production is shown on the vertical axis, where affiliate production is defined as the output of foreign plants of horizontal and vertical multinationals (i.e., the output of a horizontal firm's home plant is not counted). In the center of the box where the two countries are identical, exactly half of all world X output is affiliate output: all firms are horizontal multinationals and their home and foreign plants are identical, so half of each firm's output is affiliate output.

Figure 16.3

But affiliate sales can be even higher and this occurs in Figure 16.3 when one country is both small and skilled-labor abundant. In this case most or even all firms are vertical multinationals with their headquarters in the small, skilled-labor-abundant country. But this in turn means that all (or virtually all) plants and all production are in the other (large) country, and so all of world output is classified as affiliate output and sales.

Figures 16.4 and 16.5 help sharpen our intuition by showing two restricted versions of the model. Figure 16.4 eliminates the possibility of vertical firms (the first plant must be located with the headquarters) and so only national firms and horizontal multinationals can exist. The consequence is we should observe multinationals when the countries are similar in size and in relative endowments. Figure 16.5 does the opposite, ruling out horizontal multinationals by eliminating firm-level scale economies: the fixed costs of a two-plant firm are double the fixed costs of a single-plant firm. In this case, the is no multinational activity between identical countries and activity is maximized when one country is both small and skilled-labor abundant.

Figure 16.4 Figure 16.5

A nice thing about Figures 16.3-16.5 is that they provide clearly testable hypotheses, and further hypotheses follow from the size of trade costs relative to firm and plant-level scale economies. We will discuss this later in the chapter, but the consistent empirical finding is that the data give far more support to the horizontal model. The purely vertical case in Figure 16.5 is overwhelming rejected and the world looks much more like Figure 16.4 than 16.5: multinational activity is concentrated among similar, high-income countries. Multinational activity and investment from high income to low income countries is of much less importance, especially relative to what seems to be the popular impression.

16.6 Outsourcing versus internalization (vertical integration)

The final node in the decision tree shown in Figure 16.1 concerns the choice to maintain ownership of a foreign production facility or to outsource / license a foreign firm to produce for the multinational. While this is an old question in the international business literature, it has been only in the last two decades that it has attracted interests from international trade economists. There are two principal approaches and we will try to present simple versions here.

Some of the first formal models of the internalization decision were published in the late 1980's and 1990's, and draw their empirical motivation from the strong association of multinationality with knowledge-based assets such as those described in the previous paragraph (see Caves 2007 and Markusen 2002). On the one hand, these assets (or the services thereof) are easily transferred overseas, such as providing a blueprint, chemical formula, or procedure to a foreign plant. On the other hand, the same characteristics that make it easy to transfer these assets makes them easily learned by foreign managers, agents, or licensees. Once the agent sees the blueprint or formula he or she could defect to produce the product in a new firm. Knowledge is non-excludable, at least after some period of time.

About the same time as this last set of papers appeared, an important advancement to the internalization question was being developed by several authors, and we will review and explain a key paper by Antrás (2003). All of these authors substitute the term outsourcing for the converse term internalization. Their approach is sometimes termed the “property-rights” approach to the firm. The new literature combines a number of separate elements that together produce a coherent model that offers clear empirical predictions. The first element (in no particular order) is the assumption that production requires “relation-specific investments”, meaning that a multinational and a foreign individual or firm must incur sunk investments prior to production that have no outside value if the relationship breaks down. The second element is the assumption of incomplete contracting: certain things are simply not contractible or alternatively any contract on these items is not enforceable. The assumptions of sunk investments and non-contractibility lead to a third problem, which is ex-post “hold up”. What happens after production occurs cannot be contracted ex ante, and so each party has some ability to negotiate ex post and to prevent the other party from fully utilizing the output.

The first set of papers, focusing on the non-excludability of knowledge which is learned by a local agent or manager, can be explained by a simple version of Markusen (2002). To economize on things, we will deal only on the issue of foreign production versus exports in order to focus clearly on the role of non-excludability of knowledge.

(1) The MNE introduces (or attempts to introduce) a new product every second time periods. Two periods are referred to as a "product cycle". A product is economically obsolete at the end of the second period (end of the product cycle).

(2) The probability of the MNE successfully developing a new product in the next cycle is $1/(1+r)$ if there is a product in the current cycle, zero otherwise (i.e., once the firm fails to develop a new product, it is out of the game). The probability of having a product in the third cycle is $1/(1+r)^2$ etc. Ignore discounting.

(3) The MNE can serve a foreign market by exporting, or by creating a subsidiary to produce in the foreign market. Because of the costs of exporting, producing in the foreign country generates the most potential rents.

(5) But any local manager learns the technology in the first period of a cycle and can quit (defect) to start a rival firm in the second period. Similarly, the MNE can defect, dismissing the manager and hiring a new one in the second period. The (defecting) manager can only imitate, but cannot innovate and thus cannot compete in the next product cycle.

(6) No binding contracts can be written to prevent either partner from undertaking such a defection. I will assume that the MNE either offers a self-enforcing contract or exports.

R- Total per period licensing rents from the foreign country.

E- Total per period exporting rents ($E < R$).

F- Fixed cost of transferring the technology to a foreign partner. These include physical capital costs, training of the local manager, etc.

T- Training costs of a new manager that the MNE incurs if it dismisses the first one (i.e., if the MNE defects).

G- Fixed cost that the manager must incur if he/she defects. This could include costs of physical capital, etc.

L_i- Licensing or royalty fee charged to the subsidiary in period i (i = 1,2).

V Rents earned by the manager in one product cycle: $V = (R-L_1) + (R-L_2)$.

V/r- Present value of rents to the manager of maintaining the relationship.

The manager ("a" for agent) has an individual rationality constraint (IR): the manager must earn non-negative rents. The manager also has an incentive-compatibility constraint (IC): the manager must not want to defect in the second period: second-period earnings plus the present value of earning from future products (if any) must exceed the single-period one-shot return from defecting.

$$(R - L_1) + (R - L_2) \geq 0 \qquad \text{IR}_a \qquad (16.13)$$

$$(R- L_2) + V/r \geq (R - G) \qquad IC_a \qquad (16.14)$$

where $V = (R- L_1) + (R - L_2)$ is the present value to the manager of the future rents, if there are any. $(R - G)$ is the payoff to unilaterally defecting.

The MNE similarly has an individual rationality constraint (IR): the MNE must earn profits at least equal to the profits from exporting. The MNE also has an incentive-compatibility constraint: the MNE must not want to defect (fire the manager) in the second period.

$$L_1 + L_2 - F \geq 2E \qquad IR_m \qquad (16.15)$$

$$L_2 \geq R - T \qquad IC_m \qquad (16.16)$$

Combine the IC constraints.

$$R - T \leq L_2 \leq G + V/r \qquad (16.17)$$

Firm's objective is to minimize V subject to this incentive compatibility. Making V as small as possible subject to (16.17) holding gives us:

$$2R - L_1 - L_2 = V = r(R - T - G) \geq 0 \qquad \text{(rent share to the manager)} \qquad (16.18)$$

Our first results is then that, if $R \leq G + T$, the MNE captures all rents in a product cycle, henceforth referred to as a rent-capture (RC) contract, and the agent's IR_a constraint holds with equality. This occurs when

(1) The market is relatively small.
(2) Defection costs for the MNE (T) are high.
(3) Defection costs for the manager (G) are high.

If $R > T + G$, there is no single-product fee schedule that will not cause one party to defect. In this case, the manager's IR_a constraint does not hold as a strict equality: that is, the MNE shares rents with the manager and the amount of rent sharing is given in (16.18). This is a credible commitment to a long-term relationship that we could think of as a subsidiary. However, it is costly for the multinational and if it gets too costly then the multinational will choose exporting instead: dissipating some rent is preferable to sharing a larger total. This is the inefficiency caused by the lack of contractibility of knowledge, and may lead (will lead form many parameter values) the firm to make a welfare-inefficient choice to export and dissipate total surplus rather than share a larger surplus with the local agent.

As noted above, the “property-right” approach works rather differently. Here I present a much simplified version on Antrás (2003). The idea is that the firm and the local agent must make ex ante investments that are not contractible. The multinational invests capital K and the agent invests labor L. Ex post, they divide the surplus via the Nash bargaining solution, with the firm getting share s and the agent the share (1-s). We cannot go through an analysis of Nash bargaining here, but many readers may know from industrial-organization or game theory that the equilibrium share is a function of each party’s “bargaining power” and its outside option. Antrás assumes the firm has a bargaining power parameter of at least 1/2. Ownership is defined as a property right to anything left (e.g., an intermediate input) in the event of bargaining breakdown.

Under outsourcing, denoted with the subscript "o", any intermediate output produced is worthless to both in the event of a breakdown. Thus the outside option of both the firm and the agent is zero even though the agent owns what is left. Under FDI, denoted with the subscript "v" for vertical integration, the multinational has some use for what is left and so has an outside option. If $\phi > 1/2$ denotes the multinational's bargaining power and δ denotes the share of the total potential rent (under a successful contract) the Nash solution gives the multinational the following shares.

$$s_o = \phi \qquad s_v = \delta + \phi(1 - \delta) \qquad s_v > s_o > 1/2 \tag{16.19}$$

The second equation is a common result in the bargaining literature: the firm gets its outside option, plus its bargaining share of the total minus the sum of the outside options (the agent's outside option is zero).

Let the prices of K and L equal one. Profits from the project are given by

$$\Pi = \left(\frac{K}{\beta}\right)^{\beta}\left(\frac{L}{\gamma}\right)^{\gamma} - K - L \qquad \beta + \gamma < 1 \tag{16.20}$$

Knowing that there will be holdup and ex post bargaining with equilibrium share *s*, the firm and the agent respectively maximize the following in choosing the input they control.

$$\max_{K} \; s\left(\frac{K}{\beta}\right)^{\beta}\left(\frac{L}{\gamma}\right)^{\gamma} - K \qquad \text{Firm chooses } K \tag{16.21}$$

$$\max_{L} \; (1 - s)\left(\frac{K}{\beta}\right)^{\beta}\left(\frac{L}{\gamma}\right)^{\gamma} - L \qquad \text{Agent chooses } L \tag{16.22}$$

The first-order conditions for capital (chose by the firm) and labor (chosen by the agent) are respectively given by

$$\left(\frac{K}{\beta}\right)^{\beta - 1}\left(\frac{L}{\gamma}\right)^{\gamma} = \frac{1}{s} = (1 + t_k) \qquad t_k \equiv \frac{(1 - s)}{s} \tag{16.23}$$

$$\left(\frac{K}{\beta}\right)^{\beta}\left(\frac{L}{\gamma}\right)^{\gamma - 1} = \frac{1}{1 - s} = (1 + t_l) \qquad t_l \equiv \frac{s}{(1 - s)} \tag{16.24}$$

The last equality in each line is just to emphasize that the agent problem here is much like having a tax on capital of $t_k = (1\text{-}s)/s$ and a tax on labor of $t_l = s/(1\text{-}s)$. Indeed, the first-order conditions are exactly those of a single integrated firm maximizing profits subject to these input taxes. Ex post holdup is like each party being able to tax the other. Given (16.19), vertical integration is effectively a lower tax on capital than outsourcing, and vertical integration is effectively a higher tax on labor. A first-best outcome (i.e., by an integrated single decision maker) would be to set both equations to one (the true prices of capital and labor) rather than to something greater than one. Rearranging (16.23) and (16.24) we get

$$(K/\beta)^{1 - \beta} = s(L/\gamma)^{\gamma} < (L/\gamma)^{\gamma} \qquad \text{firm's choice of } K \text{ for a given } L \tag{16.25}$$

$$(L/\gamma)^{1-\gamma} = (1-s)(K/\beta)^{\beta} < (K/\beta)^{\beta} \qquad \text{agent's choice of } L \text{ for a given } K \tag{16.26}$$

where the quantities on the far right of each expression are the first-best outcomes.

Antrás notes that (16.25) and (16.26) can be thought of as reaction or best-response functions for the firm and agent respectively and the situation is shown in Figure 16.6.[1] When s = 1 and (1-s) = 1 in the two equations respectively, we have the first-best outcome in which the two reaction functions F* (firm) and S* (agent) cross at point A. Both outsourcing and vertical integration shift in both reaction functions. However, since $s_v > s_o$, vertical integration shifts in F less (to F_v) than outsourcing does (to F_o) in Figure 16.6. Conversely, outsourcing shifts in S less (to S_o) than vertical integration does (to S_v) in Figure 16.6. The vertical integration equilibrium is at point B in Figure 16.6 while outsourcing leaves us at C. Both outcomes are inefficient, but vertical integration has a relatively less under utilization of capital while outsourcing has a relatively less under utilization of labor.

Figure 16.6

It may therefore be intuitive that a firm in a capital-intensive industry will choose vertical integration while a firm in a labor-intensive industry will choose outsourcing, but this requires a fair bit more algebra to prove. The two first-order conditions can be solved to yield

$$\frac{K}{L} = \frac{s}{1-s}\frac{\beta}{\gamma} \qquad \frac{L}{\gamma} = \frac{1-s}{s}\frac{K}{\beta} \tag{16.27}$$

These can be substituted back into the first order conditions to give the equilibrium inputs

$$K = \beta\left[s^{1-\gamma}(1-s)^{\gamma}\right]^{\frac{1}{1-\beta-\gamma}} \tag{16.28}$$

$$L = \gamma\left[s^{1-\beta}(1-s)^{\beta}\right]^{\frac{1}{1-\beta-\gamma}} \tag{16.29}$$

The profit level for the firm in (16.20) is then given by:

$$\Pi = (1-\beta)\left[s^{1-\gamma}(1-s)^{\gamma}\right]^{\frac{1}{1-\beta-\gamma}} \tag{16.30}$$

The choice between vertical integration or outsourcing then reduces to evaluating the ratio

$$\frac{\Pi_v}{\Pi_o} = \left[\frac{s_v^{1-\gamma}(1-s_v)^{\gamma}}{s_o^{1-\gamma}(1-s_o)^{\gamma}}\right]^{\frac{1}{1-\beta-\gamma}} \tag{16.31}$$

Given the assumption that $s_v > s_o \geq 1/2$, it is true that $s_v(1-s_v) < s_o(1-s_o)$: the function $s(1-s)$ reaches a maximum value of 1/4 at *s* = 1/2 and falls off as *s* grows larger (or shrinks for that matter). As the share on labor becomes small (the industry is very labor intensive), (6.35) reduces to

$$\gamma \Rightarrow 0 \qquad \frac{\Pi_v}{\Pi_o} \Rightarrow \frac{s_v}{s_o} > 1 \tag{16.32}$$

and so vertical integration is chosen by a capital-intensive firm. At a value of $\gamma = 1/2$, implying a value of $\beta < 1/2$, we have

$$\gamma = \frac{1}{2} \qquad \frac{\Pi_v}{\Pi_o} = \left[\frac{s_v(1 - s_v)}{s_o(1 - s_o)}\right]^{\frac{0.5}{1-\beta-\gamma}} < 1 \tag{16.33}$$

Thus outsourcing is chosen by a labor-intensive firm. Antrás then presents empirical evidence, using intra-firm versus arm's-length trade, that gives good support to his model.

16.7 Summary

A number of empirical observations, particularly the fact that the high-income countries are both the major sources but also the major recipients of multinational investment has led economists to re-think the old idea of foreign direct investment as simply capital flows, moving from where capital is abundant to where it is scarce (developing countries). A cornerstone of the new theory is the existence of firm-level (or multi-plant) economies of scale arising from the joint input or non-rivaled property of knowledge based assets. Blueprints, processes, managerial techniques and so forth are costly to create but once created can be applied to multiple plants at very low additional cost.

Combined with trade costs and differing factor intensities across activities (e.g., skilled-labor intensive headquarter services versus less skilled-labor-intensive production), knowledge-based firm scale economies generate rich models and predictions about what sort of firms will arise between what sorts of country pairs. Further, the predictions of the models are empirically testable and have generated good support in formal econometric estimation.

In parallel with this new learning is a literature on outsourcing versus vertical integration, previously termed internalization (the converse of outsourcing). In one strand of literature, the same properties of knowledge-based assets that lead firms to expand abroad in the first place create difficulties in contracting and in securing property rights over the knowledge. This can lead firm to choose vertical integration (subsidiaries) or exporting when they would in principal like to choose outsourcing (licensing) on the basis of cost. A more recent approach focuses on what is referred to as the property-rights theory of the firm and to the existence of hold-up problems arising from relation-specific investments. In the approach reviewed here (Antrás), capital-intensive firms choose vertical integration while labor-intensive firms choose outsourcing. Empirical work is in its early stages, but results are supportive to the theory.

REFERENCES

Antrás, Pol (2003), "Firms, Contracts, and Trade Structure", *Quarterly Journal of Economics* 118(4), 1375-1418.

Braconier, Henrik, Pehr-Johan Norbäck, and Dieter Urban (2005), "Reconciling the Evidence on the Knowledge-Capital Model." *Review of International Economics* 13(4): 770-86.

Brainard, S. Lael (1997), "An Empirical Assessment of the Proximity-Concentration Tradeoff between Multinational Sales and Trade", *American Economic Review* 87(4): 520-544.

Carr, David L., James R. Markusen, and Keith E. Maskus (2001), "Estimating the Knowledge-Capital Model of the Multinational Enterprise." *American Economic Review* 91(3): 693-708.

Caves, Richard E. (2007). *Multinational Enterprise and Economic Analysis*. Cambridge: Cambridge University Press, third edition.

Davies, Ronald B (2008), "Hunting High and Low for Vertical FDI". *Review of International Economics*, 16(2), 250–267, 2008

Dunning, John H. (1973). "The Determinants of International Production". *Oxford Economic Papers* 25: 289-336.

Ethier, Wilfred J. and James R. Markusen (1996), "Multinational Firms, Technology Diffusion and Trade," *Journal of International Economics* 41, 1-28.

Helpman, Elhanan (1984), "A Simple Theory of Trade with Multinational Corporations", *Journal of Political Economy* 92(3): 451-471.

Markusen, James R (1984), "Multinationals, Multi-Plant Economies, and the Gains from Trade", *Journal of International Economics* 16(3-4): 205-226.

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

Markusen, James R. and Anthony J. Venables. (2000) "The Theory of Endowment, Intra-Industry, and Multinational Trade", *Journal of International Economics* 52(2): 209-234.

ENDNOTES

1. The curvature of the two reaction functions follow from the assumption that $\beta + \gamma < 1$, which in turn implies that $\gamma/(1-\beta) < 1$ and $\beta/(1-\gamma) < 1$.

Table 16.1: World statistics, 2007

Affiliates sales as a share of world exports	1.82
Value added of affiliates as a share of world GDP	0.11
Affiliate exports as a share of world exports	0.33
Affiliate exports as a share of affiliate sales	0.18

Source: UNCTAD World Investment Report

Table 16.2: Developed countries as source *and* destination for FDI: developed countries' share of world totals

	FDI inflows	FDI outflows
2007	0.66	0.85
	FDI inward stock	FDI outward stock
1990	0.73	0.92
2007	0.69	0.84

Source: UNCTAD World Investment Report

Table 16.3: Local sales, export sales, and imports of foreign affiliates, 2007

	Affiliate local sales as as share of total sales	Affiliate exports as a share of total sales	Affiliate imports as a share of affiliate sales
United States	0.72	0.28	0.06
Japan	0.65	0.35	0.43
Sweden	0.78	0.22	0.16

UNCTAD World Investment Report 2008, Annex Tables B.12, B.15, B16

Table 16.4 Sales by US and Swedish manufacturing affiliates: shares in total, 2003

	local sales	export sales back to parent country	export sales to third countries
USA 2003	0.60	0.13	0.26
Sweden 1998	0.65	0.08	0.27

Source: Markusen (2002), Davies, Norbaeck, Tekin-Koru (2009)

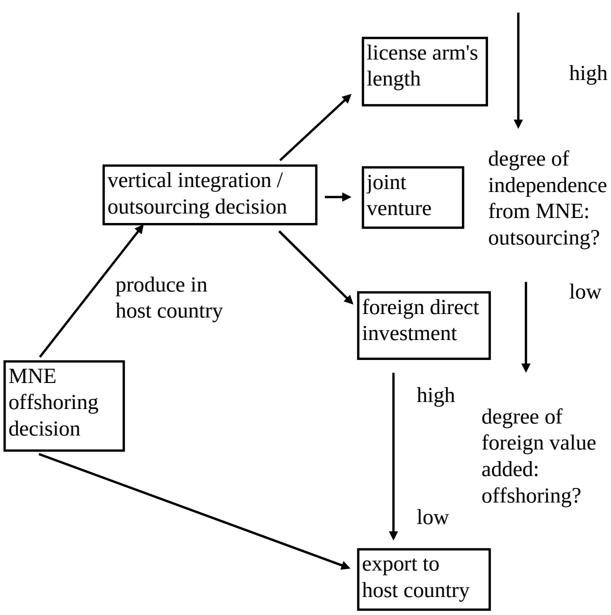


Figure 16.1: Decision tree for FDI

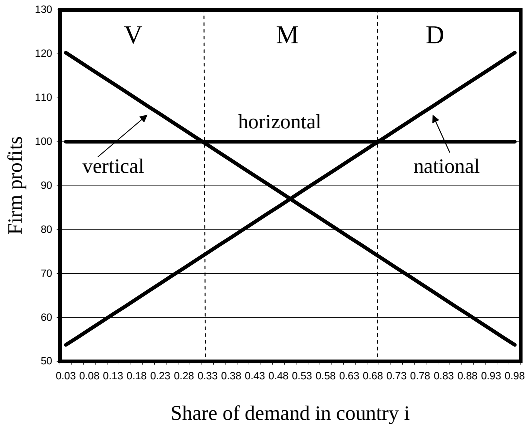


Figure 16.2: Relative size differences and choice of regime,

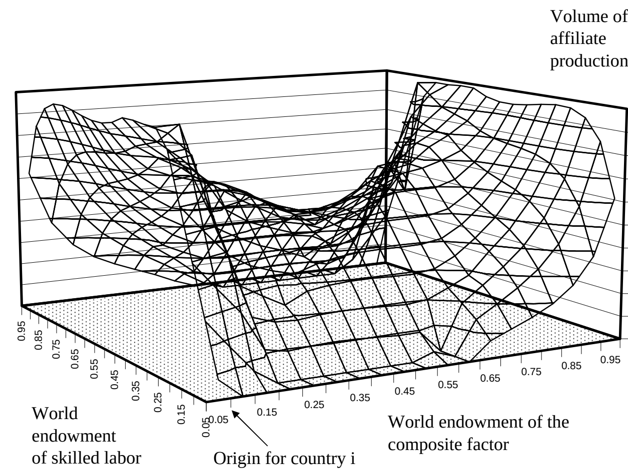


Figure 16.3: Affliate sales in the knowledge-capital model

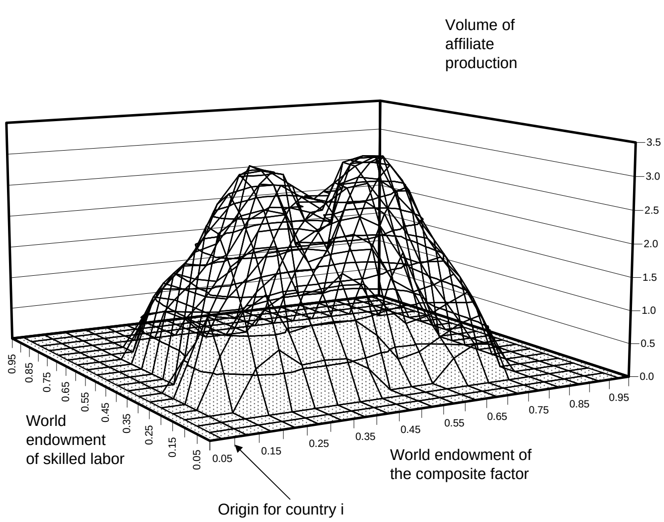


Figure 16.4: Affliate sales in the knowledge-capital model, restricted to horizontal firms

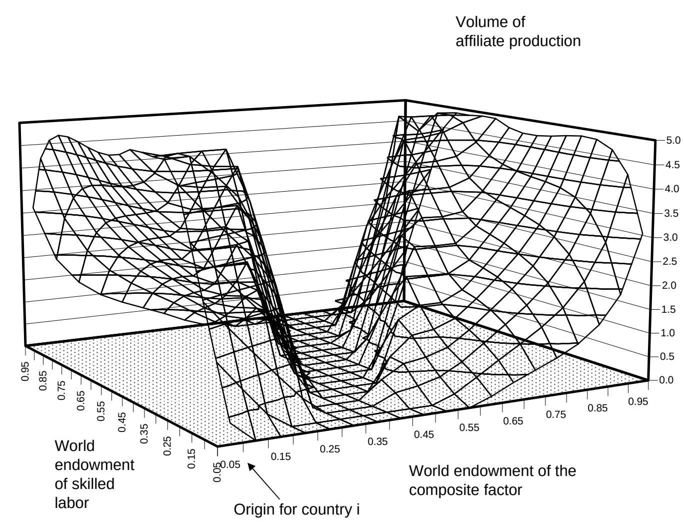
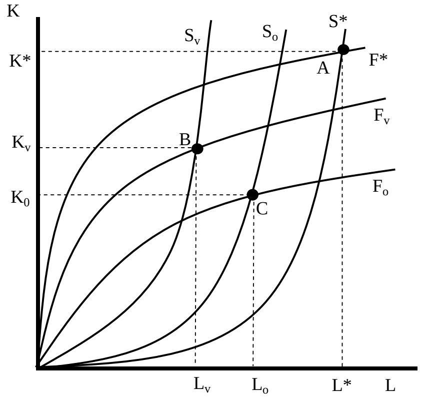


Figure 16.5: Affliate sales in the knowledge-capital model, restricted to vertical firms



A - first best, B - vertical integration, C - outsourcing

Figure 16.6: vertical integration versus outsourcing

Chapter 17

FRAGMENTATION, OFFSHORING, AND TRADE IN SERVICES

17.1 Stylized facts, basic concepts

There are a few topics that have attracted a good deal of attention but which do not fit neatly into any simple model of trade. One of these is the fact that more things seem to have become traded in the last decade or two, including intermediate goods and various types of services. Indeed, twenty years ago many trade economists more or less defined services as non-traded.

A couple of terminology points first. There is a distinction between new goods and newly-traded goods. The former of course refer to things that simply did not exist before while the latter refer to things that existed as non-traded goods or services which have now become traded. The latter is the topic of this chapter. The breaking up of the production process for a final good into different stages which are physically located in different countries has been referred to by a number of names, but we will tend to use the term “fragmentation”. Vertical specialization is another name, though this has been used in a more narrow sense. The introduction of newly traded goods is sometimes referred to as the expansion of trade at the “extensive” margin whereas more trade in the same goods is often called expansion at the “intensive” margin. By whatever name, trade economists have argued persuasively that much of the great increase in trade over the last decades has come by expansion at the extensive margin (Hummels, Rapoport, and Yi 1998; Hummels, Ishii, and Yi 2001; Yi 2003).

Often intermediate goods are produced in one or several countries while final assembly occurs in another country. Of course, final assembly may also occur in many countries which is the stuff of horizontal multinationals discussed in the previous chapter. In many cases such as in electronics, the “upstream” intermediates or components are more capital and skilled-labor intensive while “downstream” assembly is less-skilled-labor intensive but there are surely examples which run the other way. Finally, we again draw the distinction between “outsourcing”, which we reserve for production outside instead of inside the ownership boundaries of the firm and “offshoring” which we reserve for geographical location of production outside the home country.

There are obvious advantages that fragmentation offer to the world economy. Think back to Chapter 15 where we used the concept of the factor-price-equalization set. If the countries’ endowments are outside this set, then trade in goods alone cannot exhaust all gains from trade. In that Chapter, we discussed how added gains from trade can be captured by allowing trade in factors. The same is true in the presence of increasing returns to scale as we discussed in that Chapter. Trading intermediate goods and services accomplishes the same thing and in the previous chapter we advanced the view that multinational firms could be thought of primarily as vehicles for the intra-firm trade in services such as management, technology, marketing, and finance.

There is no general theory of fragmentation at this time and it is probably fair to say that the literature consists of a lot of fairly specific situations. Along with this, there few if any general welfare results about fragmentation and the expansion of trade at the extensive margin. Specifically, it is impossible to predict that all countries will be made better off by added traded goods, never mind the further complication of factor-price changes within economies. The fundamental problem is that newly-traded goods and services introduce general-equilibrium effects that mean that some individual countries could be made worse off through adverse price changes. We can give an example of this which is taken from Markusen and Venables (2007). Suppose that our country is really good at making shirts, a product produced in two stages: fabric (textiles) and cut-and-sew. Our country is not great at either of those individually, but good enough at each to be very competitive on the world market. Now fragmentation of the industry is made possible and suppose that we find that some countries are really good at textiles and some others really good at cut-and-sew: neither group was good at the other activity initially which is why we were so competitive. With trade in intermediates allowed, the world price of shirts is going to fall and our country is going to experience an adverse terms-of-trade change. In spite of our best efforts to adapt, we could be worse off.

The next two sections of the Chapter will look at quite specific examples of fragmentation from two recent papers. The first is from Markusen and Venables (2007) and Markusen (2010) while the second is from Grossman and Rossi-Hansberg (2008). The two provide an interesting contrast. Section 4 provides a somewhat general gains-from-trade theorem with respect to the effects of newly-traded goods and services and we do indeed see the crucial role of changes in the prices of existing traded goods for establishing welfare results. The final theory section looks at the issue of trade and FDI in services and the so-called four modes of trade.

17.2 Fragmentation and newly-traded intermediate goods

The first special case we will consider might be termed “conventional”, with no normative connotation that this is good or bad, in that it follows quite a number of more general formulations in the literature. Examples are Deardorff (2001, 2008), Jones and Kierzkowski, (2001), and a series of articles in Arndt and Kierzkowski (2001). Empirical analysis is found in Feenstra and Hanson (1996, 1997). Initially-traded final goods are assembled from intermediate goods and services or “components” and may also use primary factors directly. Innovations in transportation or management allow some of the components to be directly traded.

An example is shown in Figure 17.1 which sticks to the Heckscher-Ohlin tradition for familiarity: there are two final goods, X_1 and X_2 and two primary factors K and L (or these could be skilled and unskilled labor)There are three intermediate goods, denoted A, B, and C which are produced from primary factors L and K. A is the most capital intensive, B is in the middle, and C is the most labor intensive. X_1 is costlessly assembled (no added primary factor requirements) from A and B and X_2 is costlessly assembled from B and C. All production functions as assumed to be Cobb-Douglas so that the value shares of factors in intermediates and value shares of intermediates in final production are constants. Figure 17.1 gives these shares. Including intermediate use, X_1 is capital intensive overall with factor shares 65 for capital and 35 for labor in the simulation we present shortly. X_2 is the mirror imagine, with share 35 for capital and 65 for labor.

Figure 17.1

We cannot present a great many cases here and will stick with just one specific example in order to illustrate some general ideas. We assume three countries, labeled 1,2 and 3. Country 1 is capital (skilled labor) abundant with an endowment ratio K/L = 85/15 and country 3 is labor abundant with an endowment ratio K/L = 15/85. Country 2 is in the middle, with an endowment ratio 50/50. The model is very symmetric, both with respect to production and with respect to countries. Note that the country 1 and 3 endowment ratios are most extremely (differ more from one) than the factor intensities of *A* and *C* which are 80/20 and 20/80 respectively. This is important for some results, but we will not comment more on this in more detail.

Table 17.1 shows some simulations of this model. The model is benchmarked (parameters picked) so that the integrated world equilibrium we have discussed before gives each country a welfare level of one and all factor and commodity prices are one. The first column of results gives results when X_1 and X_2 are freely traded but no trade in intermediates is allowed. Because country's 1 and 3 have extremely endowment ratios, they are outside the factor-price equalization set, specialized in X_1 and X_2 respectively. Each of these countries has a high real price for its scarce factor and a low real price for its abundant factor, and each country has a welfare level (0.885) which is lower than in the integrated equilibrium. Because of the symmetry in the model, relative prices of *X* and *Y* are one, and so country 2 is indifferent to trade and enjoys the same welfare level as in the integrated world equilibrium.

Table 17.1

The second column of results in Table 17.1 allows free trade in *A* and *C*, what we could call a "symmetric" fragmentation. Countries 1 and 3 become the sole producers of *A* and *C* respectively which are better suited to their factor endowments, and country 2 becomes specialized in producing *B* only. Countries 1 and 3 have a welfare gain, but have the usual Stolper-Samuelson effect which raises the real income of the abundant factor and lowers the real income of the scarce factor. Perhaps the most unanticipated result of Table 17.1 is the large welfare gain for country 2: it is not the case that the country with the average world endowment is unaffected by the symmetric fragmentation.

The intuition for the large gain for country 2 can be explained in relatively simple supply and demand terms. In the base case, we noted that country 2 is indifferent to trading goods only and, when trade in *A* and *C* are opened up, it remains indifferent as long as prices all remain at one. But countries 1 and 3 are not at all indifferent. They are inefficient producers of *B*, and so at prices of unity they want to export *A* and *C* to country 2 and import finished X_1 and X_2 back in exchange. This drives down the prices of *A* and *C* relative to *B*, which creates a welfare gain for country 2 in the middle: country 2 imports *A* and *C* cheaply from countries 1 and 2 respectively, assembles these with local *B*, and exports finished X_1 and X_2 back to 1 and 3 respectively. In this simulation, country 2 imports *A* and *C* at the price 0.91 while the domestic price of *B*, its specialty, is 1.10. Both factors share fully in this gain for country 2.

The third column of results in Table 17.1 shows the results for an asymmetric fragmentation in which only intermediate good *C*, the most labor intensive, is traded. *A* and *B* are non-traded. Here we see a couple of interesting results. First, note that the relative prices of X_1 and X_2 are going to change: the relative prices of X_1 rises and that of X_2 falls. Country 3 specializes more in *C* and this drives down the price of *C* and hence of X_2 in general equilibrium. In this particular example, this negative terms-of-trade effect is actually strong enough to reduce the welfare of country 3 relative to the benchmark. The Stolper-Samuelson theorem applies here, and the scarce factor *K* in country 3 suffers a loss and also does so in country 2 as the theorem would predict. But the opening of trade in *C* is pure trade creation for country 2 (it did not trade initially), so the gains-from-trade theorem fully applies and we see that its welfare rises.

Perhaps the most interesting result in the third column of Table 17.1 is that not only does country 1's welfare increase, but both factors gain. The reason for this is that country 1 does not produce or import *C* either before or after fragmentation. Country 1 produces X_1 from domestically produced A and

B: its relative factor prices are pinned down by its domestic specialization in X_1. So its welfare level and the real prices of both factors rise by the same proportion.

Table 17.1 is a very specific and special case, but it shows both the opportunities and the pitfalls that are involved in expanding trade at the extensive margin. All countries can gain, but it is possible that some countries might lose due to an adverse terms-of-trade effect. We will shown in section 17.4 that this is in fact a very general proposition: a country can only lose if it suffers a deterioration in the price(s) of its initially exported good(s), otherwise it is assured of gain. Table 17.1 also shows that it is possible though not inevitable that all primary factors in a country might gain: country 2 in the second column and country 1 in the third column.

17.3 Fragmentation and trade in "tasks"

A recent paper by Grossman and Rossi-Hansberg (2008) has proposed a rather different structure which they term "tasks". A simplified version of their formulation is shown in in Figure 17.2. There are again two final goods, X_1 and X_2 and two primary factors, *L* and *K*. A task uses just a single factor, either capital or labor. But there are many labor tasks and many capital tasks. In our simplified version here, we assume that there are just two labor tasks, L_1 and L_2, and two capital tasks, K_1 and K_2. Each final good requires all labor tasks in equal amounts and it is further assumed that there is no substitution possible among these tasks; e.g., X_1 might require exactly one unit of both L_1 and L_2 with no substitution possible. The same is true for capital tasks.

Figure 17.2

Where the final goods differ is in the ratio of labor to capital tasks and hence there remains a Heckscher-Ohlin type of factor intensity. A numerical example of this is given in Figure 17.2. X_1 is assumed to be capital intensive, requiring 65 units of each of tasks K_1 and K_2 and 35 units of each of tasks L_1 and L_2 to produce 400 units of X_1 if all prices are initially normalized at unity. X_2 is the labor-intensive mirror image: 65 units of each of tasks L_1 and L_2 are required and 35 units of each of K_1 and K_2.

There are two important differences between this and the formulation in the previous section. First, a country cannot, by definition, have a factor-endowment ratio that is more extreme than some intermediate (task) that is potentially traded. Tasks use only a single primary factor and hence that there are thing that are potentially tradeable that have a more extreme factor intensity than any country (or at worse a tie if there is a country that has only one factor). Second, if only a subset of labor or capital tasks are tradeable, then free trade in the subset is quite different from simply allowing labor or capital migration as in Chapter 15. A trabeable *L* task must, for example, be combined in fixed proportions with a non-tradeable domestic *L* tasks produced by a domestic worker.

This last property is crucial in certain results that Grossman and Rossi-Hansberg emphasize in their paper and so some intuition is warranted. Suppose we have a small country h facing fixed world prices, but it is capital abundant and specialized in X_1 and so has a wage rate higher than the rest of the world. If task L_1 becomes tradeable, our country will of course import L_1. But each unit of imported L_1 must be combined with a unit of L_2 produced by a domestic worker. Let's do the thought experiment of holding domestic factor prices constant in country h. Production of X_1 is now strictly profitable since the labor "composite" input (tasks L_1 and L_2 in fixed proportions) is now cheaper: a combination of the lower cost of importing L_1 and the existing wage for task L_2. Capital is in fixed supply, so firms will demand more labor until the imported L_2 is equal to the domestic labor force, all of whom now do task L_2.

This expansion in X_1 output will raise the relative and real price of the composite capital input (in fixed supply) and lower the price of the composite labor input. However, although the final price of the *composite* labor input may be below its price before task trade, the former is a combination of the lower import price of L_1 and the higher domestic wage paid for task L_2. Thus it may well be that the final price of domestic labor is higher than before task trade and this is indeed a principal finding that Grossman and Rossi-Hansberg highlight. When a limited number of tasks can be traded, the domestic factor performing those tasks (competing with the imports) may be made better off. Grossman and Rossi-Hansberg make an interesting analogy between this trade in a limited number of tasks and technical change. In our simple case just discussed, the composite L task becomes cheaper at a given domestic wage rate, akin to a technical improvement in the production of the composite L task.

In order to provide some concreteness with the minimum possible algebra, suppose that the *composite* capital task (K_1 and K_2 in fixed proportions, perfect *complements*) and the *composite* labor task (L_1 and L_2 in fixed proportions, perfect *complements*) are perfect *substitutes* in producing final good X_1. This is equivalent to saying that there are two alternative ways of producing X1:

$$X_1 = \left[\min(K_1, K_2) + \min(L_1, L_2)\right] \quad \Rightarrow \quad X_1 = \min[K_1, K_2] \qquad X_1 = \min[L_1, L_2] \qquad (17.1)$$

Let r and w be the domestic prices of capital and labor, and t the price of imported L_1. The *fixed* world price of X_1 is denoted p_1 and superscript n denotes no trade in tasks and superscript t denotes trade in L_1 permitted. Then the cost functions corresponding to (17.1) and the price-equals-marginal-cost conditions for country h are as follows:

$$p_1 = (r + r) \qquad p_1 = (w^n + w^n) \qquad p_1 = (w^t + t) \qquad t < w^n \qquad (17.2)$$

The second equation applies in the absence of trade in tasks and the third equation with imports of task L_1. In this very special case, it is clear that allowing trade in L_1 does not affect the return to domestic capital. However, with $t < w$, it must *raise* the price of w, the wage of the domestic labor that (superficially) competes with imported task L_1: $w^t > w^n$.

While Grossman and Rossi-Hansberg put a great emphasis on this interesting result, it is of course a special case. Table 17.2 presents some simulations of our simplified version of their model. In the top panel of the Table the countries are the same size and have the same factor endowments as countries 1 and 3 of the previous section. X_1 and X_2 are assumed Cobb-Douglas as in the previous section while the two labor tasks are required in fixed proportion and the same for capital tasks. Factor shares etc. are chosen to be the same as in the previous section so that the benchmark when the countries trade in X_1 and X_2 only produces exactly the same outcome as in Table 17.1. The two models are complete equivalent when only X_1 and X_2 are traded.

Table 17.2

The first counterfactual simulation in Table 17.2 (second column of results) allows trade in tasks L_1 and K_2, a symmetric fragmentation similar to that in the previous section where we allowed A and C to both be traded. Both countries gain and the Stolper-Samuelson property again appears: there are real gains to each country's abundant factor and real losses for each country's scarce factor. The second counterfactual allows only trade in L_1, with the results reported in the third column of Table 17.2. Both countries gain, but country h, the importer of task L_1 gains less. Lower down we see the reason: the added production capacity in country h and lower capacity in country f moves the terms of trade against country h: the relative price of X_1, its export good, declines. The direction of factor-price changes is the same as in the symmetric fragmentation though the magnitudes are different.

Panel B in the lower part of Table 17.2 conducts the experiment we discussed above: country h is small (everything is the same except the factor endowment of country f is multiplied by 100). Country h's welfare and the prices of both factors are higher in the benchmark now because the relative price of X_1, country h's export good, is higher in the benchmark. The first counterfactual reported in the second column of Panel B is again the symmetric fragmentation of allowing trade in both L_1 and K_1. Country h gets a huge welfare increase, but yet again we see that the scarce factor labor suffers a real income loss. Capital exports from country h reduce the capital available for local production and so lower the marginal product of domestic labor in spite of being able to import L1. The third column of Table 17.2 shows the result emphasized above in which allowing trade in L_1 only actually raises the prices of both factors of production in country h. It also demonstrates that the result does not rely on the perfect substitutes assumption used to explain the result above (Table 17.2 assumes that L and K composites are Cobb-Douglas substitutes exactly the same as Table 17.1).

As a final point, both Tables 17.1 and 17.2 report the proportional change in the volume of trade in the counterfactuals. In all cases, these numbers are positive which certainly concurs with intuition: more things traded means more trade. For exactly this reason, we might note that Markusen and Venables (2007) and Markusen (2010) emphasize that this is not a general result and that it is easy to construct cases in which some countries will trade less as a consequence of fragmentation. A simple example should suffice. Suppose that there are two goods, wheat and cars. Suppose that country h is generally really good at cars but simply cannot produce tires. Then if only final goods are traded, country h will be forced to export wheat in exchange for cars. But now allow tires to be traded. Country h can now just import the tires and pay for this by exporting a few cars and/or a little wheat. The volume of trade falls. Markusen and Venables show that this is by no means pathological and it occurs for significant subsets of countries in their multi-country simulations.

17.4 A gains-from-trade theorem

General gains-from-trade results in trade theory are hard to obtain. The basic textbook result only proves that free trade is better for a country than autarky or, somewhat more generally, any level of restricted (but not subsidized!) trade is better than autarky. However, there is no general result that says that more trade is better than less (but positive) trade. This is due to the possibility of adverse price changes; e.g., further liberalization by a large country can make it worse off. But this is exactly the type of result we are seeking here: under what conditions will the introduction of an additional set of trade possibilities to an already-existing set of traded goods improve welfare? Our approach follows the classic "revealed preference" methodology that we have used several times in the book. The theorem itself is a much simplified version of a more general proof in Markusen (2010).

X denotes a vector of final goods quantities, with the vector p denoting their prices. Similarly, Z denotes a vector of intermediate goods with prices q. In our explicit model above, X (now denoting X_1, X_2) and Z (now denoting A, B, and C) were disjoint sets. The important distinction is that (a) only X goods enter into welfare-producing consumption and (b) we will *define* the absence of fragmentation to mean that goods cannot be traded for intermediate use. That is, the *Z* goods are initially non-traded.

Superscript f denotes prices with fragmentation (trade in intermediates allowed), while subscript n denotes prices in the initial no-fragmentation equilibrium. Thus vectors of *world* final goods prices with and without fragmentation are p^f and p^n respectively. q^f denotes the world prices of the *Z* goods when they can be traded and q^n denotes their domestic prices when they are non-traded. In the simplified proof here, we assume that the *X* goods are all freely and costlessly traded (again, see Markusen (2010) for a more complete proof with trade costs). Thus the domestic prices of the *X* goods must equal world prices.

The notation that we have been using, *X* for production and *D* for consumption is now a little awkward so we make a modification in order to avoid confusion. Just *X* and *Z* will be used: a subscript "*o*" on an *X* or *Z* quantity denotes production (output) of the good and a subscript "*c*" denotes consumption. For an intermediate good, "consumption" means its use as a domestic input in a final good. Thus for both final and intermediate goods, a positive value for production minus consumption of the good indicates it is an export good, and so forth. We will also simplify the notation a bit: rather than keep writing lots of summations, we will use a shorthand such as

$$\sum p_i X_i \equiv pX$$

We assume a competitive undistorted economy, and so production efficiency applies and the value-added of the economy in regime j (j = f, n) must be maximized at that regime's prices. Specifically, the value added in regime f must be greater than or equal to regime-n value added when the latter is evaluated at regime-f prices. Total value added for the economy is given by the value of final goods production minus intermediate usage ("consumption" of intermediates) plus the value of intermediate goods production. The production efficiency condition is then given by

$$(p^f X_o^f - q^f Z_c^f) + q^f Z_o^f \geq (p^f X_o^n - q^f Z_c^n) + q^f Z_o^n \tag{17.3}$$

This rearranges to an inequality on final production plus net exports of intermediates.

$$p^f X_o^f + (q^f Z_o^f - q^f Z_c^f) \geq p^f X_o^n + (q^f Z_o^n - q^f Z_c^n) \tag{17.4}$$

The last term in brackets on the right-hand side is zero: intermediates are not traded in the n-regime, so production and consumption are the same for each good.

$$p^f X_o^f + (q^f Z_o^f - q^f Z_c^f) \geq p^f X_o^n \tag{17.5}$$

Now we introduce the balance of trade constraints for each regime, which require that the value added in production equals the value of final consumption.

$$p^f X_o^f + (q^f Z_o^f - q^f Z_c^f) = p^f X_c^f \tag{17.6}$$

$$p^n X_o^n = p^n X_c^n \tag{17.7}$$

Substitute the right-hand side of (17.6) for the left-hand side of (17.5)

$$p^f X_c^f \;\geq\; p^f X_o^n \tag{17.8}$$

Now add the right-hand side of (17.7) and subtract the left-hand side, and also add and subtract the term $p^f X_c^n$. (17.8) then becomes

$$p^f X_c^f \;\geq\; p^f X_o^n + p^n X_c^n - p^n X_o^n + p^f X_c^n - p^f X_c^n \tag{17.9}$$

$$p^f X_c^f \;\geq\; p^f X_c^n + p^f (X_o^n - X_c^n) - p^n (X_o^n - X_c^n) \tag{17.10}$$

$$p^f X_c^f \;\geq\; p^f X_c^n + (p^f - p^n)(X_o^n - X_c^n) \tag{17.11}$$

The classic revealed-preference criterion for gains from fragmentation requires that the left-hand side of (17.11) is greater than or equal to the first term on the right-hand side: the value of regime-f consumption is revealed preferred to regime-n consumption valued at regime-f prices. So a *sufficient* condition for this to be true is that the second additive term on the right-hand side of (17.11) is non-negative.

Inequality (17.11) tell us that a sufficient condition for the added trade in the intermediate *Z* goods to improve welfare $(p^f X_c^f \;\geq\; p^f X_c^n)$ is that world prices of the initially-traded *X* goods don't change $(p^f = p^n)$. Actually all that is needed is that the world *relative* prices of the *X* goods don't change (a numeraire rule could be that the *X* prices sum to one, so then the price vectors p^f and p^n will be the same before and after *Z* trade). The absence of terms-of-trade changes for the *X* goods could be referred to as a "neutral" or "symmetric" fragmentation, emphasizing that these are *definitions* (e.g., a symmetric fragmentation is one that leaves relative final-goods prices unchanged).

The absence of relative price changes for initially-traded goods rules out the sort of example we posed about: a country good at shirts and the result of fragmentation is that efficient fabric and cut-and-sew producers lead to a fall in the world relative price of shirts. This welfare loss cannot happen if the relative price of shirts to other initially-traded goods don't change and in fact the country will generally benefit by exploiting some differences between its (old) domestic prices for fabric and tailoring and the new world prices.

The condition that world prices do not change is very restrictive and "overly" sufficient. Consider again the last term in (17.11): if this is non-negative, then this is sufficient for gains from fragmentation. Let $X_e^n \;\equiv\; (X_o^n - X_c^n)$ denoted the vector of net exports in regime n (no trade in intermediates), also referred to as the "initial" net export vector. The last term in (17.11) can be written as either of the following

$$(p^f - p^n)X_e^n \;\geq\; 0 \quad \textit{or} \quad p^f X_e^n \;\geq\; 0 \quad \textit{since} \quad p^n X_e^n \;=\; 0 \text{ by trade balance} \tag{17.12}$$

The first expression is just a simple correlation between the changes in domestic final-goods prices and the initial net export vector. We will have gains from fragmentation if "on average" the prices of initially-exported goods rise and the prices of initially-imported goods fall; that is, there is an "aggregate" terms-of-trade improvement. Noting that the value of the initial trade vector at initial prices is zero (trade balance: equation (7)), we have the second expression. If, after fragmentation, the country were to retain its initial net export vector, it would run a trade surplus if the inequality is strict. It could then improve its welfare by cutting some exports and/or increasing some imports to restore trade balance.

Unfortunately, the inequality in (17.12) is unlikely to hold for all countries. If there are only two countries, then it cannot hold for both, because the elements of their net export vectors are equal and opposite in sign. One country must have an aggregate terms-of-trade deterioration. However, we must emphasize again that (17.12) is a *sufficient* condition for $(p^f X_c^f \geq p^f X_c^n)$ which is itself a *sufficient* condition for gains. It is easy to produce numerical examples where a country benefits substantially from fragmentation and outsourcing in spite of a terms-of-trade loss and all countries gain.

17.5 Trade and foreign direct investment in business services

Services cover a very wide range of items, and we have made a decision here to restrict our analysis to business services which are generally intermediate goods (services supplied to other businesses), consistent with the overall focus of this chapter. It is indeed the case that trade and FDI in business service has increased greatly in the last decade or two. Statistics are presented in Table 17.3 for the United States and a very thorough analysis is found in Francois and Hoekman (2008). It is also clear that much of the increased activity is mediated by multinational firms, whether that is actual FDI (we measure sales of foreign affiliates) or intra-firm trade between parent and affiliate. Table 17.3 shows that both trade and FDI have increased somewhat faster than all trade and that they now account for a very significant proportion of US international economics activity.

One reason to study trade and FDI in business services, in addition to the rate of growth, is that they are often subject to very different restrictions than trade in goods. Barriers to trade and investment in services can be roughly broken down into what we could call "natural" economic costs and "policy-imposed" costs. For the former, we are thinking of things like communications and transport costs (workers flying between countries), language, customs, time zones, the need for face-to-face interaction and so forth.

Policy-induced barriers, henceforth "barriers" for this section, to trade in services take diverse forms and therefore affect service suppliers' cost functions differently. Regulatory policies, in addition to explicit and implicit barriers to trade in services, generally fall into one of five basic categories. First, there can be quantity-based restrictions imposed on services suppliers that explicitly restrict the volume of services imported, similar to a quota. The use of a fixed number of licenses available or access to only certain firms or sectors also falls into this category. If a "quota" type policy is only applied to imported services, then we would expect to see more multinationals establishing affiliates in the market (all other things equal). This is similar to a "tariff jumping" activity discussed in traditional theories of the multinational firm.

Thirdly, there are numerous barriers to establishment that restrict foreign supply of services due to the high costs of establishing a commercial presence. Policies regarding licensing procedures, requirements and fees can be prohibitive. Bureaucratic red tape, requirements for local management, or lack of transparency all have detrimental effects on the fixed costs of establishing commercial presence for multinationals. These may create a substitution effect with multinationals supplying the service from abroad rather than establish an affiliate.

Fourthly, barriers to trade and establishing commercial presence in services may take the form of restricting the use of inputs. This category can include restrictions on workers, required percentages of locally produced material inputs, as well as barriers or limits on the use of networks or media for promotion and/or marketing purposes. These policies can greatly increase the costs of operations for foreign suppliers and may be prohibitive to entering the market.

The last category of restrictions encompasses the various domestic regulatory barriers that take many forms and are often overlooked when discussing impediments to trade and investment. These include policies regulating professional qualification, residency and citizenship restrictions, obligatory membership in local professional association, juridical requirements, and limitations of inter-professional cooperation. While the policies and regulations may not explicitly target foreign firms they often have this effect in practice. Regulations on professional qualifications are important domestic policies to have so as to guarantee a level of skill and professionalism to consumers. However, when these policies require residency, citizenship, or involve re-certification for professionals with comparable certifications from another country, they become costly. While they are not explicit barriers to trade these policies severely increase the fixed costs (time and money) for a firm establishing commercial presence and may be prohibitive as well.

It is beyond the scope of this chapter to present a detailed analysis of these issues, but we would like readers to at least be aware of the basic ideas involved. Figure 17.3 presents a schematic of the general principals. Final consumption is assumed to be composed of manufacturing and agriculture. Business services are an intermediate input into both of these industries which also use primary factors of production skilled and unskilled labor. Following closely on our modeling of multinational firms in the previous Chapter, business services can be decomposed into a headquarters activity and offices. Banks, finance, and insurance companies, for example, may have a headquarters in New York or London, and have offices throughout the world. Similar to horizontal multi-national (which is what many services multinationals are), a firm has one headquarters but possibly many offices.

Figure 17.3

Figure 17.3 notes the differences in trade in services and foreign direct investment in services. In the upper part of the chart, we note that trade in services occurs when there is an international geographic separation between the service firm and the service purchaser in agriculture or manufacturing. In the lower part of the Figure, we note that FDI in services occurs when there is a geographic separation between headquarters and one or more offices. Either type of separation can occur without the other; for example, a horizontal firm tends to have geographic separation between headquarters and some offices, but not between the offices and service buyers.

As briefly discussed above, quite a range of natural and policy restrictions impact on whether or not it is easy and/or legal to fragment headquarters from offices and fragment the service provider from final user. Figure 17.4 characterizes these in terms of what are know as the four modes of service trade, which comes from the World Trade Organization's GATS: general agreement on trade in services. The schematic in Figure 17.4 has two countries, north and south, and the headquarters of a service firm headquartered in the north is represented by the central box, and a possible northern office and southern office are represented by the two boxes to either side.

Figure 17.4

The Figure shows four possible ways in which services can be provided by the offices to north or south manufacturing and agricultural firms. The channel on the upper left has the headquarters and office both in the north and providing services to northern firms. We mark this as "always feasible" since it is a strictly domestic activity. Mode 1 of the GATS is defined as cross-border trade in services, which is the channel on the lower left of Figure 17.4: the northern office of the northern firm supplies services to southern firms. The northern firm may also have a southern office, which is by definition FDI in services. This is known as Mode 3 trade in services in the GATS terminology. This southern office can supply services to local firms, which is represented by the channel on the lower right of Figure 17.3. Finally, the southern office may be able to supply services to northern firms which would require both Modes 1 and 3 being feasible. An example of this last type of supply is call centers which are located in developing countries such as India. A customer in the north needing service from a northern firm finds their phone call re-routed to a call center in India.

For completeness, GATS mode 2 is when the buyer travels to the seller's country of origin. Traditionally, this category has been dominated by tourism. But the extensive margin of Mode 2 is expanding as in "medical tourism". Patients in high-income countries are finding that they can get high-quality surgery or dental work done in places as diverse as Hungary, India, and Mexico for a fraction of the cost of having it done at home. We have seen websites in Ireland, for example, that advertises dental holidays in Hungary, in which you get round-trip plane tickets, two days in a resort, and all your dental work done for less than just the latter would cost in Dublin. GATS Mode 4 is the international movement of person to work abroad, typically for limited periods of time rather than permanent migration. There is obviously some overlap and fuzziness between this and Mode 3. If Gregorz the Polish plumber goes to work in Denmark, that is Mode 4; if Gregorz Plumbing Incorporated opens a one-man office in Copenhagen, that is apparently Mode 3.

As suggested earlier, both natural and policy-imposed restrictions can make it costly or even prohibitive for a firm to engage in one or more of the modes shown in Figure 17.4. If face-to-face contact is required, for example, then Mode 1 can be prohibitively costly. If countries restrict the right of foreign firms to establish or acquire local subsidiaries (known as right of establishment) then Mode 3 is difficult. The principle of national treatment means that foreign firms are suppose to be treated exactly the same as domestic firms, but this is often violated in practice (de facto) by rules such as requiring managers to be local citizens and so forth.

Finally, arguments about whether or not something is a barrier and how significant is it abound between countries. Part of the difficulty is that many service industries are by their very nature highly regulated. Consider banking, finance, insurance, architecture and construction, telecommunications, legal services and medical services. These tend to be regulated in all countries and while foreign firms see such regulations as barriers, the local firms and governments always tend to claim that they are non-discriminatory against foreign firms.

17.6 Summary

It has been documented that a good deal of the expansion of trade over the last decade or two has been through the extensive margin, meaning trade in goods and services which were not previously traded, rather than through the intensive margin: trading larger quantities of the same things. Much of this new trade seems to be in intermediate goods and services as firms fragment the production process to take advantage of favorable factor prices or other considerations across geographic locations.

The theoretical literature on fragmentation and outsourcing is now extensive, but there is no general model we can rely on. We presented two special cases from recent literature that posit different mechanisms as to how we might think about fragmentation and newly-traded goods. These are interesting but even in these rather restrictive special cases results still depend on choices of parameters such as country sizes and factor intensities.

A simplified gains-from-trade theorem pinpoints one difficult in trying to obtain general results. This difficulty is that, when new goods and services become traded, the prices of existing goods change in general-equilibrium which leads to terms-of-trade effects for the countries involved. A country could lose through fragmentation because the components and stages of production for something the country was very good at can now be individually done better by several other countries. It is easy to create specific examples in which everyone gains but unfortunately also not impossible to create examples in which someone loses.

We closed with a discussion of trade and FDI in intermediate business services, a fast-growing component of total world economic activity. The theoretical issues here are much the same as in the previous Chapter on multinational firms and the choice between exporting and FDI. What make trade and FDI in business services of particular interest is that many of the natural and policy barriers to trade and investment differ substantially from those restricting trade in goods.

REFERENCES

Arndt, Sven W. and Henryk Kierzkowski (editors) (2001), *Fragmentation: New Production Patterns in the World Economy*. Oxford University Press, Oxford.

Brainard, S. Lael. and Susan. Collins (eds.) (2006), *Brookings Trade Forum 2005: Offshoring White-collarWork* (Washington, DC: Brookings Institution).

Deardorff, Alan V. (2001), "Fragmentation across Cones," in Sven W. Arndt and Henryk Kierzkowski, eds., *Fragmentation: New Production Patterns in the World Economy*, Oxford: Oxford University Press, 35-51.

Deardorff, Alan V. (2008), "Gains from Trade and Fragmentation", in Steven Brakman and Harry Garretsen, eds., *Foreign Direct Investment and the Multinational Enterprise*, CESifo Seminar Series, Cambridge: MIT Press, 155-169.

Feenstra, Robert C. and Gordon H. Hanson (1996), "Globalization, Outsourcing, and Wage Inequality", *American Economic Review*, 86, 240–45.

Feenstra, Robert C. and Gordon H. Hanson (1997), "Foreign Direct Investment and Relative Wages: Evidencefrom Mexico's Maquiladoras", *Journal of International Economics*, 42, 371–93.

Francois, Joseph and Bernard Hoekman (2008), "Liberalization of Trade in Services: A Survey", World Bank working paper (under revision for the *Journal of Economic Literature*).

Grossman, Gene. M., Estaban Rossi-Hansberg (2008), "Trading Tasks: A Simple Theory of Offshoring", *American Economic Review* 98, 1978-1997.

Hanson, Gordon. H., Raymond J. Mataloni, and Matthew J. Slaughter (2005), "Vertical production networks in multinational firms", *Review of Economics and Statistics* 87, 664-678.

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

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

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

Leamer, Edward and James Levinsohn (1995), "International trade theory; the evidence", in: Gene Grossman and Kenneth Rogoff (editors), *Handbook of International Economics, Vol. 3*. North-Holland, Amsterdam, 1339–1394.

Markusen, James R. (2010), "Fragmentation and offshoring: a general gains-from-trade theorem and some specific cases", working paper.

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

Markusen, James R. and Bridget Strand (2009), "Adapting the Knowledge-Capital Model of the Multinational Enterprise to Trade and Investment in Business Services", *World Economy* 32, 6-29.

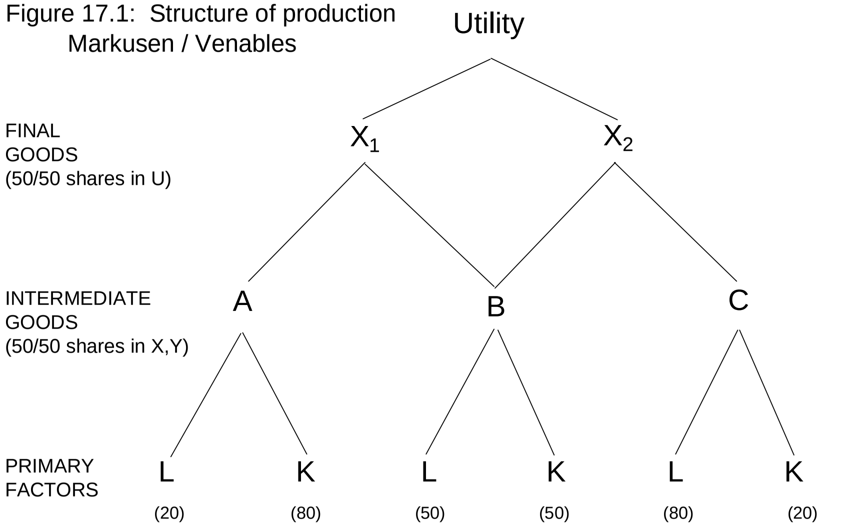
Ng, F., Yeats, A. (1999), "Production sharing in East Asia; who does what for whom and why", World Bank Policy Research Working Paper 2197, Washington DC.

Venables, Anthony J. (1999), "Fragmentation and multinational production", *European Economic Review*,43, 935-945.

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

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

Figure 17.1: Structure of production
Markusen / Venables



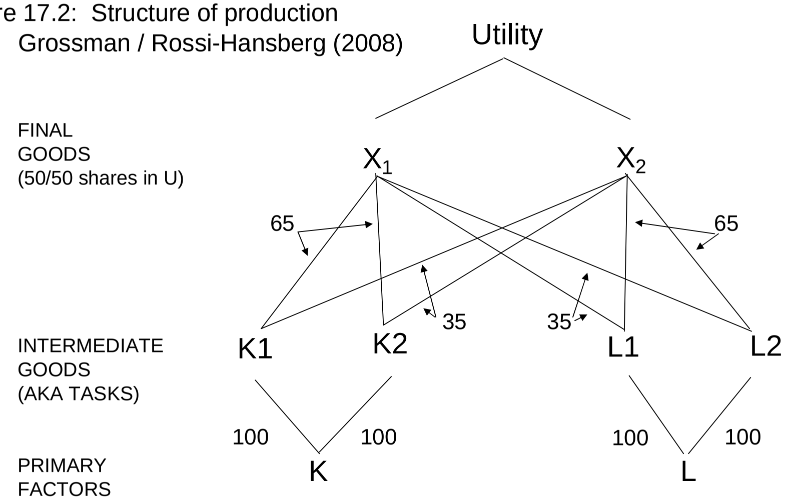
FINAL GOODS	ALWAYS TRADABLE AT COUNTRY-SPECIFIC TRADE COST
INTERMEDIATE GOODS	NONE / SOME / ALL TRADABLE AT COUNTRY-SPECIFIC TRADE COST
PRIMARY FACTORS	NOT TRADABLE

Table 17.1: Simulation results
Markusen (2010)

	Benchmark No trade in Intermediates	Trade in A (most K-int) C (most L-int)		Trade in C (most L-int) No trade in A, B	
	Level	Level	Prop change from bench	Level	Prop change from bench
Welfare country 1 (K abundant)	0.885	0.903	0.020	0.944	0.067
real price of labor	2.064	1.560	-0.244	2.202	0.067
real price of capital	0.676	0.789	0.167	0.722	0.067
Welfare country 2	1.000	1.103	0.103	1.035	0.035
real price of labor	1.000	1.103	0.103	0.894	-0.106
real price of capital	1.000	1.103	0.103	1.177	0.177
Welfare country 3 (L abundant)	0.885	0.903	0.020	0.849	-0.041
real price of labor	0.676	0.789	0.167	0.750	0.109
real price of capital	2.064	1.560	-0.244	1.420	-0.312
Price of X	1.000	1.000		1.068	
Price of Y	1.000	1.000		0.937	
Proportional change in trade volume			1.492		0.861

3.096 3.050

Figure 17.2: Structure of production
Grossman / Rossi-Hansberg (2008)



FINAL GOODS	ALWAYS TRADABLE
INTERMEDIATE GOODS	NONE / SOME / ALL TRADABLE
PRIMARY FACTORS	NOT TRADABLE

Table 17.2: Simulation results
Grossman / Rossi-Hansberg

	Benchmark No trade in Intermediates	Trade in L1 and K1		Trade in L1	
Panel A: Countries symmetric	Level	Level	Prop change from bench	Level	Prop change from bench
Welfare country 1 (K abundant)	0.885	0.994	0.123	0.923	0.043
real price of labor	2.065	1.395	-0.324	1.838	-0.110
real price of capital	0.677	0.923	0.363	0.762	0.126
Welfare country 2 (L abundant)	0.885	0.994	0.123	0.964	0.089
real price of labor	0.677	0.923	0.363	0.786	0.161
real price of capital	2.065	1.395	-0.324	1.974	-0.044
Price of X	1.000	1.000		0.922	
Price of Y	1.000	1.000		1.085	
Proportional change in trade volume			0.089		0.436
Panel B: Country h small (same relative endowments as panel A:)	Level	Level	Prop change from bench	Level	Prop change from bench
Welfare country 1 (K abundant)	1.146	2.033	0.774	1.396	0.218
real price of labor	2.674	0.433	-0.838	2.983	0.116
real price of capital	0.877	2.315	1.640	1.116	0.273
Proportional change in trade volume			0.329		0.507
	3.049			2.673	

TABLE 17.3: US Foreign Affiliate Sales and Cross Border Trade

Outward US Foreign Affiliate Sales-all countries

	1999	2005	% change
Total Sales-All Industries	2316654.8	3276024.4	41.41%
Total Private Services	353200.0	528000.0	49.49%
Information	63236.5	97069.9	53.50%
Finance & Insurance	86337.1	140341.6	62.55%
Finance	32330.4	43847.0	35.62%
Insurance	54006.6	96495.7	78.67%
PST	65290.2	97490.9	49.32%

Sales in Total Private Services = 15.25% and 16.12% of All Industries Sales in 1999 and 2005 respectively

PST- Professional, Scientific, and Technical Services

All data are in millions of 2000 US dollars

Inward Foreign Affiliate Sales-all countries

	1999	2005	% change
Total Sales-All Industries	1831561.1	2213172.5	20.84%
Total Private Services	293500.0	389000.0	32.54%
Information	46440.3	48138.5	3.66%
Finance & Insurance	95840.7	104308.3	8.84%
Finance	15651.8	25458.9	62.66%
Insurance	80188.9	78849.4	-1.67%
PST	15757.0	49648.7	215.09%

Sales in Total Private Services = 16.02% and 17.57% of All Industries Sales in 1999 and 2005 respectively

PST- Professional, Scientific, and Technical Services

All data are in millions of 2000 US dollars

Cross-Border Trade- All countries-Exports

	1999	2005	% change
All Industries	1287247.6	1586285.5	23.23%
All Industries-Affiliated	194697.1	186463.5	-4.23%
All Industries-Unaffiliated	1092550.5	1399822.0	28.12%
Total Private Services	265100.0	368000.0	38.82%
Total Private Services-Affiliated	32952.4	44441.2	34.86%
Total Private Services-Unaffiliated	73245.5	101278.7	38.27%
Financial Total	17789.3	31626.3	77.78%
Financial-Affiliated	4087.2	4312.0	5.50%
Financial-Unaffiliated	13702.2	27314.3	99.34%
Insurance Total*	3119.2	6011.5	92.73%
BPT Total	54683.1	73911.2	35.16%
BPT-Affiliated	26379.5	37062.1	40.50%
BPT-Unaffiliated	28303.5	36849.1	30.19%

Exports of total private services = 20.59% and 23.19% of trade in all industries in 1999 and 2005 respectively

*Insurance transactions are considered unaffiliated by BEA

BPT-Business, Professional, and Technical Services

Cross-Border Trade- All countries-Imports

	1999	2005	% change
All Industries	1543920.8	2177244.7	41.02%
All Industries-Affiliated	186215.3	231946.0	24.56%
All Industries-Unaffiliated	1357705.5	1945298.8	43.28%
Total Private Services	183000.0	282000.0	54.10%
Total Private Services-Affiliated	25670.2	35340.6	37.67%
Total Private Services-Unaffiliated	31048.8	53285.4	71.62%
Financial Total	9623.2	11105.6	15.40%
Financial-Affiliated	6130.7	5192.0	-15.31%
Financial-Unaffiliated	3492.5	5913.6	69.32%
Insurance Total*	9593.9	25064.2	161.25%
BPT Total	28238.2	42913.2	51.97%
BPT-Affiliated	19462.0	29868.1	53.47%
BPT-Unaffiliated	8776.1	13045.1	48.64%

Imports of total private services = 11.85% and 12.95% of trade in all industries in 1999 and 2005 respectively

*Insurance transactions are considered unaffiliated by BEA

BPT-Business, Professional, and Technical Services

Figure 17.3: Structure of production
Trade and FDI in intermediate business services

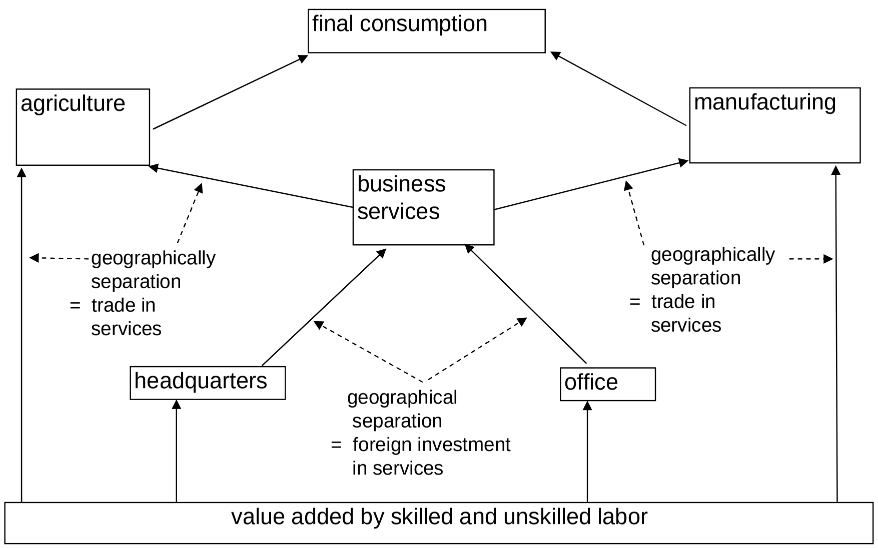
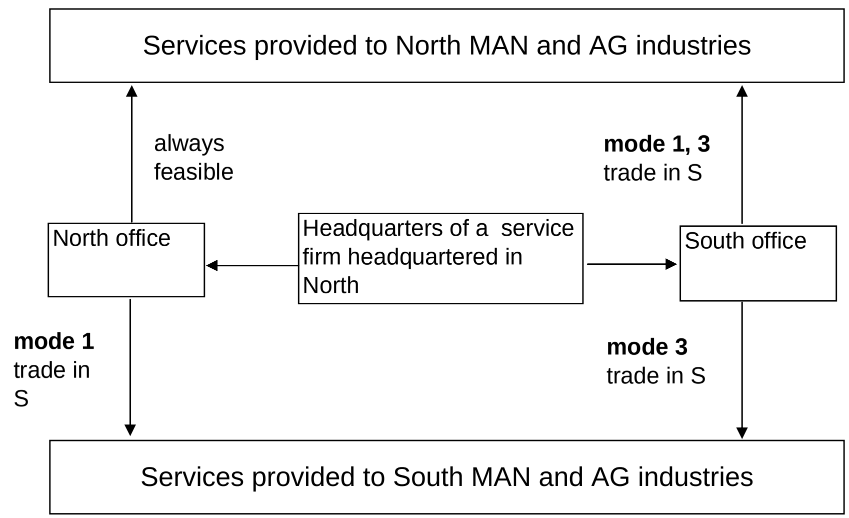


Figure 17.4: Modes of trade in services for a northern service firm



Chapter 10

DISTORTIONS AND EXTERNALITIES AS DETERMINANTS OF TRADE

10.1 Departures from our stylized world

Chapters 7, 8, and 9 discussed in some detail two of the five determinants of trade described in Chapter 6, namely, production function (technology) differences, and endowment differences. In this chapter we will discuss domestic distortions as a determinant of trade. The term distortion refers to departures from our highly stylized world which is characterized by (1) all agents are perfectly competitive and have no market power, which in turn generally implies that firms have constant returns to scale as we will discuss in the next chapter; (2) prices and quantities are freely determined by the forces of supply and demand; and (3) all interactions are though markets, meaning that there are no externalities. An externality in turn is an effect one agent (firm or consumer) has on another which is not priced through a market. Pollution is a classic example of a negative externalities, also called a spillover, in which the actions of one agent harm another and the latter is not compensated for this. A knowledge spillover in which the innovation of one firm is copied by other firms without the former being compensated is a classic example of a positive externality. Externalities, spillovers and other distortions are also often called "market failures".

The principal focus of the chapter is to understand how distortions and externalities by themselves can be causes of trade and whether or not that distortion-induced trade leads to gains from trade. The chapter retains the focus on the positive theory of trade and does not focus on normative issues such as what is optimal government policy in the presence of market failures; some of that will be considered later in the book. Words like distortion and market failure inevitably sound like bad things and they are relative to a best possible outcome. But keep in mind that the terms refer both to failures in which too much of something is produced such as pollution and in which too little of something is produced, such as knowledge.

There are two distinct parts of the chapter. The first part focuses on taxes and subsidies. The examination of taxes and subsidies is intended as an example of the effects that a wide range of government policies can have on trade. For example, environmental policies and regulations impact on firms costs, and hence have effects on outputs and trade similar to those produced by taxes. We are not asserting that commodity and factor taxes rank with factor endowments as a cause of trade, but we do believe that collectively, government policies have a much more profound impact on trade than is suggested in most international trade textbooks. One theme of the chapter is that government policies can generate trade, but it not necessarily beneficial trade.

As before, the approach will be to neutralize other factors so that a clear understanding of the specific effects of each can be obtained. Thus throughout the tax analysis, we will assume that either (A) we have a single country facing fixed world prices, or (B) there are two countries that are identical in all respects. They have identical technologies, identical factor endowments, identical homogeneous utility functions, and constant returns and perfect competition in production. In the absence of the distortions that we introduce, the two countries would have no incentive to trade, or alternatively, the free trade equilibrium would be identical to autarky.

10.2 Distinguishing among consumer, producer and world prices

When we introduce taxes and subsidies into the analysis, it becomes important to distinguish prices paid by consumers from prices received by producers. Once we introduce trade, consumer and producer prices must be distinguished from world prices, the prices at which the country can trade. Throughout this Chapter, we will use the notation q to represent consumer prices, p to represent producer prices, and p^* to represent world prices. This notation will also be used in later Chapters, such as the

Chapter on tariffs.

In order to focus on trade issues, we will also make the assumption throughout the Chapter that there is no government sector per se; the government returns all tax collections to consumers in lump-sum fashion and/or raises all subsidies by lump-sum taxation. We implicitly assume a very large number of consumers, with each consumer getting a check or a bill which gives to (or takes from) the consumer his or her share of taxes (subsidies). Each consumer regards their bill or check as being unaffected by their own purchases. For examples, if a consumer pays $1 in sales tax, the consumer gets a refund of only $1/N of that amount where N is the number of consumers (consider the refund if there are 100 million consumers). Thus each consumer does indeed regard the tax as raising prices, even though the tax is return to all consumers collectively. Similar comments apply to subsidies.

We will specify taxes and subsidies in an ad valorem (percentage of value) form throughout the Chapter rather than in specific form. t will denote a tax and s a subsidy. *Ad valorem* taxes are quoted as rates. A sales tax of 5%, for example, would mean a tax rate of $t = .05$ in this context. (*Specific* taxes, on the other hand, are quoted in monetary units per unit of the good: the US gasoline tax is quoted in cents per gallon.) Thus the relationship between consumer and producer prices with a tax or subsidy is given as follows.

$$\begin{aligned} q &= p(1+t) > p \qquad \textit{tax} \\ q &= p(1-s) < p \qquad \textit{subsidy} \end{aligned} \tag{10.1}$$

A tax raises the consumer price above the producer price while a subsidy lowers the consumer price below the producer price. A tax rate $t = .05$ raises the consumer price 5% above the producer price: $q = p(1.05)$. When there are only two goods, the effects of a tax on one good are equivalent to a subsidy on the other good. In order to see this, consider the commodity price ratios resulting from a tax on X_1 versus a subsidy to X_2.

$$\begin{aligned} \frac{q_1}{q_2} &= \frac{p_1(1+t)}{p_2} > \frac{p_1}{p_2} \qquad \textit{tax on } X_1 \\ \frac{q_1}{q_2} &= \frac{p_1}{p_2(1-s)} > \frac{p_1}{p_2} \qquad \textit{subsidy on } X_2 \end{aligned} \tag{10.2}$$

We see from (10.2) that a subsidy to X_2 and a tax on X_1 induce the same "wedge" between the consumer and producer price ratios.

Figure 10.1 gives autarky equilibrium at point E when there is *either* a tax on X_1 or a subsidy on X_2. In order to focus on the distortions per se, we will assume throughout this Chapter that the tax revenue is redistributed lump sum and the subsidy is raised by a lump sum tax. These latter assumptions are reflected in the fact the consumption and production bundles are the same even though, in the case of a tax, for example, the consumption bundle costs more than the value of those goods at producer prices. The consumers pay more than the producers receive because of the tax, but then they receive an income in excess of the value of production due to the fact that they receive the tax refund. Let D denote consumption (demand) and X denote production quantities. For a tax on X_1,

$$q_1D_1 + q_2D_2 = p_1(1+t)X_1 + p_2X_2 = \left[p_1X_1 + p_2X_2\right] + \left[p_1tX_1\right] \tag{10.3}$$

The left-hand side of (10.3) is consumer expenditure at consumer prices. The first bracketed term on the right-hand side is income received from production (payments to factors of production), while the second term on the right-hand side is redistributed tax revenue. Thus consumer expenditure equals consumer income. A similar analysis of a subsidy simply requires us to change the sign of *t*.

Figure 10.1

Figure 10.1 illustrates the distortionary effect of the tax on X_1 or subsidy on X_2. Welfare is lower at E that at the undistorted competitive equilibrium at A. The producer price ratio *p* is tangent to the production frontier TT' while the consumer price ratio *q* is tangent to the indifference curve through E. The tax causes the consumers to perceive *Y* as more costly than it actual is to produce, or a subsidy to *X* causes consumers to perceive *X* as relatively cheaper than what it actually costs to produce.

The previous paragraph should not be taken to suggest that all taxes or subsidies are bad. First, governments usually raise revenues in order to provide public goods that are not or cannot be provided by markets. No account is taken of public goods in this analysis. Second, not all taxes are distortionary or as distortionary as the commodity tax shown here. For example, in the present model, an equal ad valorem tax on both goods would leave the relative consumer and producer prices equal. Such a *set* of taxes is non-distortionary.

Finally, some government policies are imposed to correct an *existing* distortion in the economy, such as an environmental externality. In such a situation, Figure 10.1 might accurately depict the effects of a pollution tax (on X_1) on production and trade, but the indifference curves no longer indicate welfare change. Welfare may be improving due to lower pollution (i.e., there is actually a third good, environmental quality, not shown in the diagram). More will be said about taxes in the presence of existing distortions later in the book.

10.3 Taxes and subsidies as determinants of trade: a small open economy

Suppose that the home country faces fixed world prices. Assume also that these prices just happen to be equal to home's autarky price ratio such that home does not choose to trade at these prices. This is completely unlikely, but we are simply following the strategy outlined in Chapter 6: "neutralize" all causes of trade except the one which we wish to examine. The situation is shown in Figures 10.2 where the autarky equilibrium A is also the free-trade equilibrium at price ratio p^*.

Once we introduce trade, we not only have to keep track of consumer and producer prices, but world prices as well. This in turn means that we have to specify whether a tax or subsidy is assessed on consumption or production. In the closed economy, it does not matter since production and consumption of each good are equal. But with trade, consumption and production are in general not equal, so it matters which one we are taxing. In this section, we will limit ourselves to looking only at production taxes and subsidies. These are not necessarily more common than consumption taxes, but space constraints limit the range of distortions we can deal with here. It is also true that focusing on production taxes helps build some intuition for the analysis of imperfect competition which begins in the next Chapter.

Consider a tax on the *production* of X_1 or a subsidy on the production of X_2. In this case, consumers face world prices, not producers. Consumer prices and world prices will be equal to one another, but not to world prices. The relationships among the three price ratios are given by

$$\frac{p_1(1+t)}{p_2} = \frac{q_1}{q_2} = \frac{p_1^*}{p_2^*} > \frac{p_1}{p_2} \qquad (10.4)$$

The relationships in (10.4) are shown in Figure 10.2. The producer price ratio is now greater than the consumer and world price ratios, so production is shown as taking place at point X^t in Figure 10.2. Consumption must take place along the world price ratio through X^t, and consumers now face world prices. Thus the consumption point is given by the tangency between an indifference curve and the price line p^* through point X^t. We show the consumption point as D^t in Figure 10.3. The production tax discourages production of X^1 and leads to a substitution in production toward good X^2.

Figure 10.2

Several important results are shown in Figure 10.2. First, it clearly demonstrates that government policies such as taxes and subsidies can generate trade. However, it shows equally clearly that *trade induced by the introduction of distortions is not beneficial trade*. In Figure 10.2, the country receives a welfare loss as a consequence of distortion-induced trade. Point X^* in Figure 10.2 would be the undistorted equilibrium yielding a utility level of U^*. The distorted equilibrium yields a utility level of U^t.

This is a very important results insofar as governments sometimes decide that it would be a good thing if the country produced and exported more of a certain good (e.g., "high tech" goods). We could think of Figure 10.2 as a situation generated by a government deciding that it must be good to produce and export X_2. By putting on a subsidy to the production of X_2 we do indeed get exports of X_2 and the government congratulates itself on the success of its project. However, exports generated by distortions are welfare reducing (put differently, the initial level of exports, zero, is optimal).

Welfare is reduced by this distortion, because producers make *privately* efficient choices given the prices they face, but they do not make efficient *social* choices when the do not face the true costs of producing the commodities. But again, we should separate this welfare result from the results concerning consumption and trade, since not all commodity taxes need be welfare reducing. Most countries have gasoline (petrol) taxes, for example, which have the beneficial effects of reducing pollution and traffic congestion.

10.4 Taxes and subsidies as determinants of trade: two identical countries.

Now we introduce a second country and explicitly return to our concept of the two- country, no-trade model. Suppose that we have two identical countries, both with production frontiers as in Figure 10.1, such that the point A in Figure 10.1 represents both the free-trade and autarky equilibria for both of the countries. Now country h imposes a production tax on X_1 or a production subsidy X_2. At the initial free-trade price ratio p^*, country f will wish to continue to produce and consume at A in Figure 10.3 (the same as A in Figure 10.1) while country h will wish to shift production away from X_1 toward X_2.

This cannot be an equilibrium because there will be excess demand for X_1 and excess supply of X_2 at the initial prices. The price ratio p^* must rise, giving us a new equilibrium as shown in Figure 10.3. The fall in p^* induces country h to export X_2 and import X_1, producing at X^h and consuming at D^h in Figure 10.3. Country f produces at X^f and consumes at D^f. We see that a production tax on X_1 or subsidy on X_2 can indeed generate exports of X_2, but this is not welfare improving trade. Country h creates trade by its tax or subsidy, but it is not beneficial trade.

The interesting additional result that we get from Figure 10.3 is that *country f is made better off* be h's tax or subsidy. Recall from Chapter 5 that the ability of a country to trade at any prices other than its autarky prices can make it better off (and *will* make it better off if it has no distortions). The institution of the tax or subsidy in country h now allows country f an opportunity to trade at prices different than its autarky prices. This might also help us understand why country h has to be worse off. With the countries absolutely identical, there are no opportunities for mutual gains from trade. If the distortion makes country f better off, it must make country h worse off.

Figure 10.3

The implication here is that country f should be happy when it's trading partner subsidies its exports to h. Intuitively, the subsidizing country h is selling for less than the cost of production to the benefit of the passive country f. Happiness in f is rarely the reaction in practice however to a trading partner's subsidy. In some cases, a government may simply misunderstand this gift. But our result here is not general. In a Heckscher-Ohlin world of multiple factors of production, someone in country f is surely worse off and will understandably make a political fuss. Second, suppose that there are three countries, two of which export X_2 to the third. Then the first two, call them h and f are competitors, and a subsidy to X_2 by country h is going to drive down the price of X_2 for both countries, making country f worse off as well.

Again, we will not provide a detailed normative discussion of this here. Our purpose is simply to show that a distortion can serve as a basis for trade but also that trade generated by a distortion is not necessarily good trade.

10.5 Production externalities

As noted above, externalities are another source of "market failure" yet at the same time can imply an additional source of gains from trade. Such externalities come up in quite a number of contexts, ranging from pollution to intellectual property. In this section, we will look at positive production externalities among firms in an industry. As suggested earlier, this could result from knowledge spillovers in which the innovations of one firm are quickly copied by other firms without compensating the innovating firm. Many other cases are discussed in the literature, including the increases in the range of intermediate goods as an industry or country grows, an idea going back to Adam Smith's "the division of labor is limited by the extent of the market". One interpretation of division of labor is increases in the number of specialized intermediate goods, a topic we will return to in Chapter 12.

Suppose that there is just a single factor of production labor, L, which is in fixed supply. Good X_2 is produced with constant returns to scale by a competitive industry. Good X_1 is produced by competitive firms who perceive themselves as having constant to scale, but the productivity of their labor inputs is positively related to the overall output of the industry. These competitive firms view total industry output as constant, much as we assume that competitive firms view the industry price as constant and unaffected by their own decisions. Let X_{i1} denote the output of an individual firm in industry 1 and let X_1 denote total industry output, the sum over all i firms. The production side of our economy is given as follows:

$$X_{i1} = (X_1^{\alpha})L_{i1} \qquad X_2 = L_2 \qquad \bar{L} = \sum_i L_{i1} + L_2 \tag{10.5}$$

where $0 \leq \alpha < 1$ is an externality parameter: $\alpha = 0$ is the special case of no externality, in which case the model reduces to the Ricardian model of Chapter 7. As just noted, each individual firm i in industry X_1 views total industry output as constant. In competitive equilibrium, each firm equate the value of the marginal product of labor to the wage rate, denoted w as in the Ricardian model. Competitive equilibrium is then described by

$$p_1 X_1^{\alpha} = w \qquad p_2 = w \qquad \frac{p_1}{p_2} = \frac{1}{X_1^{\alpha}} \tag{10.6}$$

Total industry output in X_1 is given by summing the first equation in (10.5) over all i firms. We do this and then rearrange the equation to given total industry output X_1 as follows.

$$\sum_i X_{i1} = X_1 = X_1^{\alpha}\sum_i L_{i1} = X_1^{\alpha}L_1 \qquad X_1^{1-\alpha} = L_1 \qquad X_1 = L_1^{\frac{1}{1-\alpha}} \tag{10.7}$$

Since α < 1, the exponent on the right-hand equation of (10.7) is greater than one: total industry output exhibits increasing returns to scale in its total labor input. Differentiate the middle equation in (10.7) along with the equation for X_2 output, making use of the total labor supply constraint.

$$(1-\alpha)X_1^{-\alpha}dX_1 = dL_1 \qquad dX_2 = dL_2 = -dL_1 \tag{10.8}$$

Divide the first equation of (10.8) by the second and rearrange.

$$-\frac{dX_2}{dX_1} = (1-\alpha)\frac{1}{X_1^{\alpha}} \tag{10.9}$$

which is the slope of the production frontier, the marginal rate of transformation. As we noted back in Chapter 2, the production frontier is a convex function (the production set is non-convex) reflecting the increasing returns to scale in X_1: the denominator of (10.9) gets smaller as X_1 gets larger. The production frontier for our economy is shown as $\bar{X}_2\bar{X}_1$ in Figure 10.4.

Figure 10.4

Now combine (10.9) with the competitive pricing condition in (10.6). This gives us a relationship between the marginal rate of transformation and the equilibrium price ratio.

$$-\frac{dX_2}{dX_1} = (1-\alpha)\frac{p_1}{p_2} < \frac{p_1}{p_2} \tag{10.10}$$

Now we discover a second issue, in addition to the convexity issue, connected with a positive production externality. The competitive-equilibrium price ratio is not tangent to the production frontier, but rather cuts it as shown in Figure 10.4. The intuition behind this result is the fact that when an individual firm expands output a little, it confers a positive productivity effect on all other firms taken together. Thus the true or "social" marginal product of an additional worker hired is greater than the "private" marginal product of an individual firm. Or to put it the other way around, the private cost of an addition worker hired is more than the true social cost. The slope of the production frontier depends on the true social cost and so it is flatter than the price ratio, equal to private marginal cost. Competitive equilibrium is at a point like A in Figure 10.4, giving a welfare level of U^a.

While we don't want to get into a detailed normative analysis here, we should note that this is a case where an offsetting distortion could increase welfare. The first-best outcome in Figure 10.4 is at point S yielding welfare U^s. This could be achieved by a subsidy to X_1 in effect compensating for or "internalizing" the externality. If the externality is due to imperfect protection of intellectual property (the innovating firms is not compensated for benefits conferred on other firms), then added intellectual property protection will act to offset the distortion and move the country toward point S in Figure 10.4. This is an application of what is known as the theorem of the second best:

> Theorem of the second best: in the presence of one distortion, the imposition of an additional and offsetting distortion can improve welfare.

We will return to this idea later in the book.

10.6 Trade and gains from trade in the presence of production externalities

Let us abstract from the issue of the price line cutting the production frontier by assuming that there are positive externalities in both X_1 and X_2 and in the same degree α. Then there will be a term (1 - α) in both the numerator and denominator of (10.10) and these will cancel out, leaving the marginal rate of transformation equal to the price ratio. In this special case, we will have an autarky equilibrium at point $X^a = D^a$ in Figure 10.5, giving an autarky welfare level of U^a.

Figure 10.5

Now assume that there are two absolutely identical economies and let them trade. It is beyond the scope of this chapter to given a detailed analysis of the adjustment process but in short, the outcome shown in Figure 10.5 at X^a continues to be an equilibrium but it is *unstable*. A small perturbation can send the two countries off to corners, each specializing in only one of the two goods. One country could specialize in good X_2 and the other in X_1 and they could each trade half of their output for half of the other country's good, leading both counties to share a common consumption point at D^* in Figure 10.5. Here is our first instance of how there can exist gains from trade even between identical countries arising from increasing returns to scale.

The outcome shown in Figure 10.5 in which the equilibrium price ratio is exactly the cord connecting the two endpoint of the production frontier a special case requiring strong symmetry assumptions. Suppose at this price ratio, the countries do not demand X_1 and X_2 in the proportions produced, but each country wants to consume a lot of X_1 and not so much X_2. Then at the price ratio shown in Figure 10.5, there will be excess demand for X_1 and excess supply of X_2. The outcome is going to have to be as shown in Figure 10.6: the relative price of X_1 will rise to clear the market. The country specializing in X_2, call that country h, consumes at D^h and country f specializing in X_1 consumes at D^f in Figure 10.6.

Figure 10.6

Note that the outcome shown in Figure 10.6 is not the only possible one. Reversing which country is which is also an equilibrium. Thus situations in which there are production externalities can be characterized by multiple equilibria. In such a situation, we want to note that this phenomenon has many implications, for development economics in particular. In the presence of multiple equilibria, a country doesn't want to end up in the "wrong" outcome: it wants to be country f in Figure 10.6, not country h. A more detailed analysis is beyond the scope of this chapter. But in closing, refer back to Figure 10.4. Suppose one country has a free market and the other one "internalizes" the externality via intellectual property protection for example. The former country is at A while the latter one is at S in Figure 10.4. If trade opens up, the country at S has a lower price for X_1 and hence will tend to end up specializing in X_1: this country will be country f shown in Figure 10.6. Internalizing positive production externalities can give a country an advantage in a situation of multiple equilibria.

10.7 Summary: what you should know

This chapter has turned our attention away from underlying production differences between countries, principally differences in technologies and in factor endowments. Instead, we neutralize these differences by assuming that technologies, factor endowments and also demand are identical across two countries. This continues the methodology outlined in Chapter 6. Here we look at distortions and externalities (which are themselves a class of distortions) to see how asymmetries in distortions across countries can generate trade and may or may not generate gains from trade for each of two countries.

One class of distortions is represented here by simple production taxes and subsidies. In such a situation, it is important to keep track of different sets of prices, in particular consumer, producer, and world prices. The principal result of this analysis is that production taxes or subsidies can generate trade between otherwise identical countries, but it is a welfare-worsening trade for the taxing/subsidizing country. This carries a very important lesson for policy makers, which is that exports should never be confused with welfare. A production subsidy to a favored sector may indeed generate exports from that sector, but these exports are being sold abroad for less than the cost of production and hence are welfare worsening.

The second class of distortions addressed here are positive production externalities or spillovers, arising due to some failure on the part of firms to be able to capture returns for benefits they confer on rival firms. Examples include the lack of protection for intellectual property and innovations, and increases in productivity arising from the finer “division of labor” in a larger market. These issues will arise again in this book.

These positive production externalities imply aggregate increasing returns to scale even though individual firms have constant returns technologies. As discussed back in Chapter 2, this can imply that the production frontier is convex (the production set non-convex) or “bowed in”. This in turn can imply positive gains from trade for each of two identical countries: with each country specializing in only one sector, productivity rises and then can exist mutual gains from trade. The situation is not simple however, and we touch briefly on issues like the existence of multiple equilibria, with the alternative equilibria having very different welfare implications for the trading partners.

The next two chapters continue to look at similar, even identical economies, and analyze how imperfect competition and increasing returns to scale offer added sources of gains beyond those arising from comparative advantage linked to differences between countries.

REFERENCES

Bhagwati, J. and T.N. Srinivasan (1983), *Lectures on International Trade*, Cambridge: MIT Press, Chapters 20-23.

Grossman, Gene M. and Estaban Rossi-Hansberg (2009), “External economies and international trade redux”, Princeton University working paper.

Ethier, Wilfred J. (1979), “Internationally decreasing costs and world trade”, *Journal of International Economics* 9, 1-24.

Kemp, Murray C. (1969), *The pure theory of international trade and investment*, New Jersey: Prentice Hall.

Markusen, James R. (1900), “Micro-Foundations of External Economies”, *Canadian Journal of Economics* 23, 495-508.

Markusen, James R. and James R. Melvin (1981), “Trade, Factor Prices, and the Gains from Trade with Increasing Returns to Scale," *Canadian Journal of Economics* 14, 450-469.

Melvin, J. R. (1970). "Commodity Taxation as a Determinant of Trade." *Canadian Journal of Economics* 3, 62-78.

Melvin, J. R. (1969). “Increasing Returns to Scale as a Determinant of Trade”, *Canadian Journal of Economics* 2, 389-402.

Figure 10.1

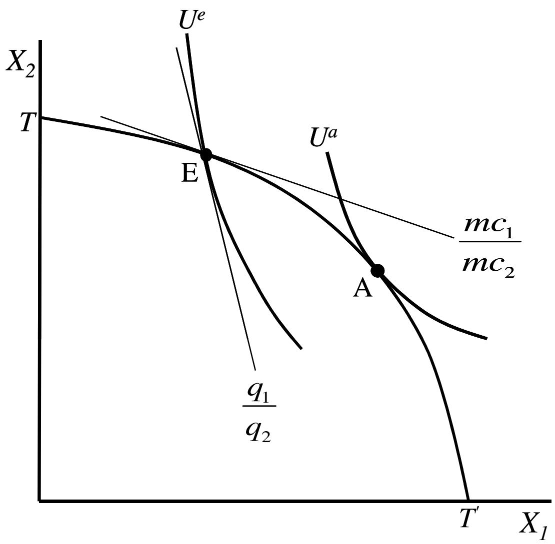


Figure 10.2

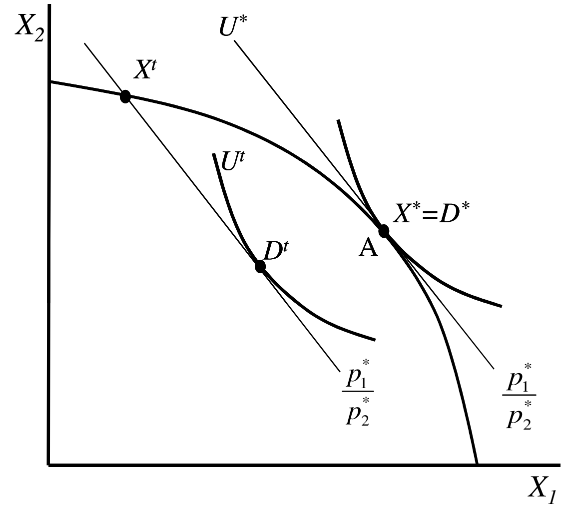


Figure 10.3

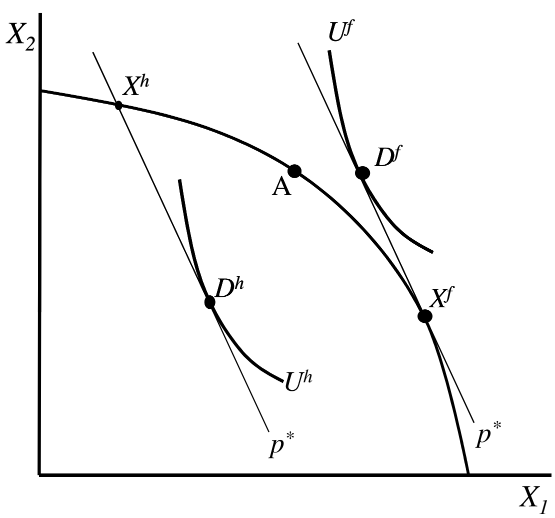


Figure 10.4

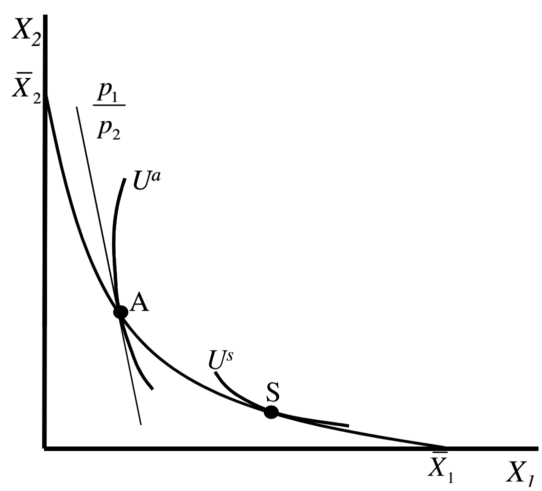


Figure 10.5

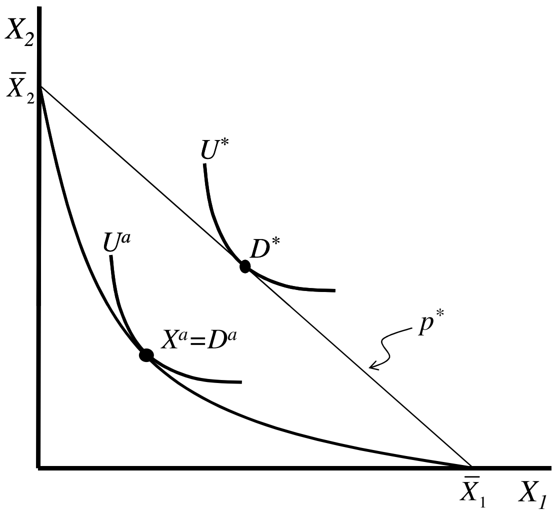


Figure 10.6