THE INDUSTRIAL INSTITUTE FOR ECONOMIC AND SOCIAL RESEARCH

WORKING PAPER No. 463, 1996

LECTURE NOTES ON INTERNATIONAL TRADE AND IMPERFECT COMPETITION

BY ANTHONY VENABLES

Revised, 18/6/96

Lecture Notes on International Trade and Imperfect competition.

Anthony Venables

London School of Economics
Houghton Street
London WC2A 2AE

Contents:

1. Introduction:

In the last fifteen years imperfect competition has come to the fore in the theoretical analysis of international trade. This development was driven by a number of factors. Researchers were motivated by the large and growing volume of international trade in similar products (intra-industry trade) between similar countries, a phenomenon not adequately explained by traditional theory based on perfect competition. They were also motivated by the fact that many of the important commercial policy questions seemed to be better answered by using tools from industrial economics than from trade theory. And on the supply side, developments in game theory and the theory of industrial organisation opened up a new set of tools to be used by researchers.

The theory that has been developed over this period succeeds in providing a fuller explanation of recent developments in trade flows, and also supports a rich set of policy analyses, linking trade and industrial policy, and illuminating economic integration. It is noteworthy that the theory can be integrated with the more traditional factor endowment and general equilibrium concerns of trade theory; in some circumstances it can be placed 'on top of' traditional theory, providing a theory of both inter- and intra-industry trade. It also opens the way to constructing a theory of 'new economic geography', incorporating the location decisions of firms and of mobile factors of production.

The objective of these notes is to provide an overview of the main developments in this literature, and a framework within which the reader may place other work. The notes are divide into two main sections. In the first our attention is on strategic interaction between firms in oligopolistic markets. The focus is on a single industry in which there are rather few firms, and the analysis can perhaps be described as 'open economy industrial economics'. We investigate the trade flows that can occur in such an environment, and look at the welfare and policy issues raised by this trade. In the second section we return to the general equilibrium concerns of traditional trade theory. To capture these a simpler model of imperfect

competition is needed -- we use the Spence-Dixit-Stiglitz framework which ignores strategic interactions, and focuses on market power derived from product differentiation. Here our primary concern is the pattern of trade, and we develop a complete synthesis of factor endowment and market access determined trade flows. We also look at factor mobility, and the possibility that this may lead to the agglomeration of economic activity.

2. Oligopolistic interaction:

If firms from different countries compete on international markets that are less than perfectly competitive, then what form does their competitive interaction take? What trade flows does it create? And what are the welfare effects of trade and of trade policy? To answer these questions we devote this section to the study of markets in which there is strategic interaction between firms.

As is often remarked, whereas there is only one theory of perfect competition, the problem with imperfect competition is that there are many theories. These theories are unified in their use of modern game theory -- we shall be analysing Nash equilibria -- but differ principally in their specification of the strategic variable chosen by firms. In the following two sub-sections we shall look first at the case where firms' strategic variables are price, and then when they are output or sales quantities. In a Nash equilibrium no firm has an incentive to change the value of its strategic variable given the values selected by other firms. However, as we shall see, the equilibrium can be quite different according as to whether prices or quantities are chosen.

The difference between price competition (Bertrand) and quantity competition (Cournot) is well known from industrial organisation. In the international trade context a further distinction becomes important. Firms may choses quantities (or prices) in each market separately, which will shall refer to as the case of segmented markets. Or they may select a single world-wide quantity (or price), which we shall refer to as integrated markets. We shall

focus on the segmented market case, merely commenting on the difference made by integrated markets.

Before undertaking formal analysis of these situations, it is convenient to specify some of the notational conventions which will be followed throughout the paper. Subscripts i, j = 1, 2... will be used to denote countries. In some contexts it is necessary to distinguish a variable both by its origin and destination country, and the first subscript will then denote origin and second destination, thus x_{12} is a quantity produced in 1 and sold in 2. If a single subscript is used then it should be clear from the context whether it refers to the location of consumption or production. We shall sometimes refer to country 1 as 'home' and country 2 as 'foreign'.

2a: Price competition:

The simplest framework we can consider is a single industry containing one firm in each country, 1 and 2. These firms have constant marginal costs, denoted c_1 and c_2 respectively, for producing a homogeneous product. We shall assume that markets are segmented, so will look at just one country's market (there is no interaction between the two markets, so the other market can be described analogously). We shall look at the country 1 market and suppose that there is a trade cost, t per unit, that the firm in 2 must incur if it is to supply country 1, making its effective marginal cost $c_2 + t$.

Each firm chooses its price, p_1 or p_2. What does the equilibrium look like, and does it involve trade? Since both firms produce the same homogeneous product, all sales will go to whichever firm has the lowest price. If D(.) is the market demand function, we therefore have demands faced by each firm,

$$\begin{aligned} &\textit{if} \quad p_1 > p_2, \quad x_1 = 0, \quad x_2 = D(p_2), \\ &\textit{if} \quad p_2 > p_1, \quad x_2 = 0, \quad x_1 = D(p_1), \\ &\textit{if} \quad p_1 = p_2, \quad x_1 = x_2 = D(p_1)/2. \end{aligned} \tag{1}$$

(the equal division of the market given in the last case is arbitrary).

Firms maximise profits, and their best response functions are illustrated on figure 2.1 by the functions $R_1(p_2)$ and $R_2(p_1)$. $p_1 = R_1(p_2)$ gives the values of p_1 that maximise firm 1's profits given the value of p_2. Its shape is determined by the fact that profit maximisation implies that firm 1 should just undercut firm 2 and thereby take the entire market, but never set price less than its unit costs, c_1, or above the monopoly price p_1^m. $R_2(p_1)$ is constructed analogously.

The Nash equilibrium is at the intersection of these best response functions, at point E. What is the pattern of trade at this point? The figure is constructed with the assumption that $c_2 + t > c_1$. The equilibrium involves prices $p_1 = c_2 + t - \delta$ (where δ is a small positive number), $p_2 = c_2 + t$, and therefore firm 1 taking the entire market -- so there is no trade. Notice however, that the possibility of trade affects the equilibrium. If trade were impossible, then the market price would be p_1^m. The possibility of imports disciplines the domestic firm, reducing its price and profits, and raising overall social welfare -- the 'pro-competitive' effect of trade.

What is optimal trade policy in this model? To answer this question suppose that t is not a real trade cost, but a tariff, creating revenue which is transferred to the country 1 government. If $c_2 \geq c_1$ then welfare is maximised by setting $t = c_1 - c_2 + \delta$, where δ is a small positive number. This value of t has two properties. First, at this value of t, $c_2 + t > c_1$; this means that the domestic (country 1) firm (by assumption the lower cost supplier) takes the entire market. Second, the domestic firm is induced to set price equal to its marginal cost, c_1, because $p_1 = c_2 + t - \delta = c_1$. Notice that this is an import *subsidy*, used to tighten the competitive discipline on the home firm; of course, the subsidy is not actually paid (there is no trade in equilibrium).

If $c_2 < c_1$ then the optimal policy is $t = c_1 - c_2 - \delta$. The foreign (country 2) firm will then take the entire market at price $p_2 = c_2 + t = c_1 - \delta$, i.e. just less than the

minimum price at which the domestic firm is willing to supply. This policy generates tariff revenue, so the unit cost of imports to economy 1 as a whole is the consumer price minus tariff revenue, $p_2 - t = c_2$, which is the world minimum production cost. The policy therefore amounts to letting the lowest cost producer supply the market (through imports), but using the tariff to extract any profits that the importer might make.

Bertrand competition with homogeneous products is notorious for giving very sharp and extreme results, and this is what we see in this case. Although quite illuminating, the model fails to predict intra-industry trade, and effects are probably too extreme to give an adequate representation of many industries. We now turn to a case where the sharp price under-cutting of Bertrand competition does not take place.

2b: Quantity competition:

We keep the same firm and country structure, but now suppose that firms use quantities -- levels of sales in each market -- as strategic variables. This is the 'reciprocal dumping' model of Brander and Krugman (1983).

Profits earned by each firm in market 1 can be expressed as:

$$\begin{aligned} \pi_1 &= P(x_1 + x_2)x_1 - c_1x_1, \\ \pi_2 &= P(x_1 + x_2)x_2 - (c_2 + t)x_2. \end{aligned} \tag{2}$$

where $(x_1 + x_2)$ is total sales in the market, and $P(x_1 + x_2)$ is the inverse demand function associated with $D(p)$.

To find the Nash equilibrium in quantities (Cournot) we look at each firm's profit maximising quantity choice given the quantity choice of other firms. Profit maximisation gives first order conditions,

$$\frac{\partial \pi_1}{\partial x_1} = P(x_1 + x_2) + x_1 \frac{\partial P}{\partial x_1} - c_1 = 0,$$

$$\frac{\partial \pi_2}{\partial x_2} = P(x_1 + x_2) + x_2 \frac{\partial P}{\partial x_2} - c_2 - t = 0, \tag{3}$$

Since the elasticity of demand, ϵ, is defined as $-\frac{1}{\epsilon} = \frac{\partial P}{\partial x} \cdot \frac{(x_1 + x_2)}{P}$, and defining market shares m_i as $m_i \equiv x_i/(x_1 + x_2)$ these reduce to the usual form of the equality of marginal revenue to cost,

$$P(x_1 + x_2)\left(1 - \frac{m_1}{\epsilon}\right) = c_1,$$

$$P(x_1 + x_2)\left(1 - \frac{m_2}{\epsilon}\right) = c_2 + t. \tag{4}$$

Equilibrium is given by values of x_1 and x_2 satisfying these equations.

How does this compare with price competition? The first point is that, even if $c_1 \neq c_2 + t$, both firms can supply the same market. Both firms receive the same price, but the higher cost firm has smaller market share (lower m_i), this raising the elasticity of its perceived demand curve (reducing m_i/ϵ) and raising marginal revenue. (There is of course a bound on the cost difference associated with both supplying the same market, which can be computed easily from the conditions above). A corollary of this is that the model supports intra-industry trade. Both firms supply both markets, even if there are trade costs associated with trade. The identical product is therefore 'cross-hauled' -- shipped in opposite directions by the two firms.

What about the welfare implications of this trade? Trade has a pro-competitive effect, reducing the price charged in each market, and this is a source of welfare gain. However, trade also reduces profits, and may redistribute them between countries. It turns out that trade

does not necessarily raise welfare, as demonstrated by the next two examples.

In this partial equilibrium framework we take as welfare criterion the sum of domestic consumers' surplus, profits accruing to the domestic firm, and domestic government revenue (if any). The effects of trade can be illustrated using standard diagrammatic techniques. In figure 2.2 the horizontal axis measures the total sales in market 1, $x_1 + x_2$, the vertical axis the price, and the inverse demand curve is given by $P(x_1 + x_2)$. Costs of both firms are equal to c, and t = 0. The autarky price is illustrated as p^a, and the price with trade is p. The autarky quantity is x_1^a; with trade firm 1's supply to this market drops to x_1, but total market supply increases to $x_1 + x_2$. (The diagram does not give marginal revenue, nor therefore does it *derive* the equilibrium prices and quantities illustrated).

The welfare effect of trade can be read off the diagram. Moving from autarky to trade increases consumer surplus by amount A + B. Profits earned in market 1 under autarky are B + C + D; with trade the home firm earns C and the foreign firm D + E. Indicating the value of foreign profits by θ and changes in going from autarky to trade by Δ we can write

$$\Delta CS + \Delta\pi = A + B + C + \theta(D + E) - (B + C + D) = A - D + \theta(D + E).$$

Profits of the foreign firm do not enter domestic welfare. However, if the two economies are symmetrical, then just as the foreign firm earns D + E in the home market, so the domestic firm earns D + E in the foreign market. We capture this by setting $\theta = 1$, this giving welfare change, $\Delta W = A + E$; this is positive, and there are gains from trade. However, suppose alternatively that the economies are not symmetrical. In particular, let the foreign economy become very small, so $\theta \to 0$. In this case the change in domestic welfare is $\Delta W = A - D$, the sign of which is ambiguous. The ambiguity is due to the 'profit shifting' effect. World welfare is increased by trade but, in asymmetric cases, profits can be shifted between countries. This may be a source of welfare loss for one of the economies.

The second example is given in figure 2.3. This is similar to figure 2.2, except that we now have trade costs, t > 0, and assume that this t is a real trade cost, e.g. a transport cost.

As before we can express changes in the domestic market as,

$$\Delta CS + \Delta \Pi = (A + B) + (C + F) + \theta(D + E) - (B + C + D + F + G)$$

where the first term is the change in consumer surplus, the second domestic profits, the third foreign profits and the fourth the loss of profits that were earned under autarky. Looking at the symmetric case, $\theta = 1$, this reduces to $\Delta W = A + E - G$ which can be positive or negative. The welfare effects are therefore ambiguous, even though the countries are identical. The reason for this ambiguity is that weighed against the pro-competitive gains from trade, are the costs incurred in cross-hauling the product. It is possible to show that with Cournot competition and homogeneous products the equilibrium involves more trade than is socially optimal. Whether trade yields higher welfare than autarky is then, as we have seen, ambiguous.

What we learn from these examples is that there is no general gains from trade theorem in oligopoly models. It has often been argued that the gains from trade are larger under imperfect competition than under perfect competition, because of pro-competitive effects. This may well usually be the case, but there is no general theorem to that effect. As we have seen, counter-examples can easily be constructed.

2c: Trade policy under imperfect competition:

In this subsection we look at both import and export policy for a country with one domestic firm and one foreign. In order to do this we first need to develop our welfare criterion more fully.[1]

For country 1 the sum of domestic consumer surplus, firms' profit and government revenue can be written as:

$$W_1 = v(p_{11}, p_{21}) + (p_{11} - c_1)x_{11} + (p_{12} - c_1 - t_1)x_{12} + \alpha t_2 x_{21} + \beta t_1 x_{12} \qquad (5)$$

The first term is the country 1 indirect utility function, measuring consumer surplus. We allow

for product differentiation, so the market 1 prices of both country 1 and country 2 firms enter this function. The second is the profits of the country 1 firm in its home market, and the third the profits of this firm on its sales in market 2. The last two terms are tariff revenue accruing to the country 1 government, where α and β take values 0 or 1 depending on the policy regime. If there is a country 1 import tariff at rate t_2, then $\alpha = 1$, transferring the tariff revenue to country 1. A country 1 export tax at rate t_1 implies $\beta = 1$.[2]

To analyse policy we shall look at small changes in the equilibrium. Totally differentiating (5) and using Roy's identity ($dv/dp_{11} = -x_{11}$, $dv/dp_{21} = -x_{21}$) we derive:

$$dW_1 = -x_{21}dp_{21} + x_{12}dp_{12} + (p_{11} - c_1)dx_{11} + (p_{12} - c_1 - t_1(1-\beta))dx_{12} + \alpha t_2 dx_{21} + \alpha x_{21}dt_2 + (\beta - 1)x_{12}dt_1 \tag{6}$$

The first two terms give terms of trade effects on imports and exports respectively. The next three are quantity changes times any wedge between marginal social benefit and marginal social cost. Marginal social benefit is price, and the marginal social cost of domestic production is cost; notice that t_1 is either a real cost or a transfer according as $\beta = 0$ or $\beta = 1$. For imports, x_{21}, the unit cost to the economy is price, and there is a wedge on import supply only if there is an import tariff, $\alpha = 1$. The final two terms are the direct effect of the policy change impacting on government revenue and, for a change in t_1, also on profits.

Import tariffs: Consider the case of an import tariff, so $dt_2 > 0$, $\alpha = 1$, $dt_1 = 0$, $\beta = 0$. If markets are segmented and marginal cost curves are flat, then a country 1 import tariff will have no effect on country 2 market variables, so $dx_{12} = dp_{12} = 0$. Equation (6) therefore reduces to

$$dW_1 = -x_{21}(dp_{21} - dt_2) + (p_{11} - c_1)dx_{11} + t_2 dx_{21} \tag{7}$$

These three effects are a terms of trade effect, (defined on the supply price), a 'firm expansion

effect' (an expansion of domestic production operating at price greater than marginal cost) and a direct revenue effect.

The terms dp_{21}/dt_2, dx_{11}/dt_2, dx_{21}/dt_2 can be found by totally differentiating the equilibrium conditions, and the expression can then be evaluated. We shall not go through the details of this, but merely make some observations. Suppose we consider a small import tariff, i.e. $dt_2 > 0$ around $t_2 = 0$. The third term on the right hand side of eqn (7) is then zero. The second is certainly positive because the tariff will expand domestic production, so creating gains from a firm expansion effect. In the first term we must look at dp_{21} and dt_2; the tariff will usually raise the price, creating welfare loss, against which must be set the tariff revenue gain. Combining these terms, what do we know about $- x_{21} (dp_{21} - dt_2)$? Notice first that, if markets are segmented then even a 'small' country might experience a change in this term; since prices are set market by market rather than on a worldwide 'integrated market' even small countries can experience terms of trade effects. What of the sign of this term? While it might be expected that the price would increase by less than the tariff -- i.e. the foreign firm absorbs some of the tariff -- this need not be so. The policy may shift the industry to a less elastic part of the demand curve, causing a large increase in p_1, and making this term negative.

Overall then, we cannot be sure about the welfare effects of a tariff. There are gains from the firm expansion effect, and even small countries may be able to change their terms of trade -- but the direction of this change is ambiguous.

Export taxes: Imposition of an export tax is a policy change $dt_1 > 0$, $\beta = 1$, $dt_2 = 0$, $\alpha = 0$. The impact of this occurs in the country 2 market so $dx_{11} = dx_{21} = dp_{21} = 0$ and equation (6) becomes,

$$dW_1 = x_{12} dp_{12} + (p_{12} - c_1) dx_{12} \tag{8}$$

This says that the policy will raise welfare if it increases the surplus that country 1 takes from

country 2 -- either by increasing price, or expanding volume where price exceeds marginal cost. Terms dp_{12}/dt_1 and dx_{12}/dt_1 can be found by totally differentiating the equilibrium conditions, and the expression can be evaluated. However, a more direct approach is more illuminating. We can express the surplus (profits plus tax revenue) country 1 derives from sales in country 2 as W_{12},

$$W_{12} = \pi_{12} + t_1 x_{12} \tag{9}$$

The profits made by firm 1 in market 2, π_{12}, can be expressed as a function of the market 2 quantities supplied and the export tax, $\pi_{12}(x_{12}, x_{22}, t_1)$. Differentiating,

$$\frac{d\pi_{12}}{dt_1} = \frac{\partial \pi_{12}}{\partial x_{12}}\frac{dx_{12}}{dt_1} + \frac{\partial \pi_{12}}{\partial x_{22}}\frac{dx_{22}}{dt_1} - \frac{\partial \pi_{12}}{\partial t_1} \tag{10}$$

In Cournot equilibrium the first term on the right hand side is zero, since x_{12} has been chosen to maximise π_{12} given x_{22}. In the final term, $\partial \pi_{12}/\partial t_1 = -x_{12}$ from the definition of profits. Differentiating (9) and using (10) we therefore have,

$$\frac{dW_{12}}{dt_1} = \frac{d\pi_{12}}{dt_1} + t_1 dx_{12} + x_{12} dt_1 = \frac{\partial \pi_{12}}{\partial x_{22}}\frac{dx_{22}}{dt_1} + t_1 dx_{12} \tag{11}$$

To evaluate the effects of a small export tax (i.e. around $t_1 = 0$) we need only know $\partial \pi_{12}/\partial x_{22}$ and dx_{22}/dt_1. The former is certainly negative -- an increase in rival's output reduces profits; and the latter positive -- an export tax raises the rival's output. A small export tax therefore reduces welfare and, conversely, a small export subsidy raises welfare.

This is the celebrated Brander-Spencer (1985) result, which has attracted so much attention. In a competitive model subsidising export sales would always reduce welfare. Here it has the opposite effect, essentially because the government, by committing to a subsidy, is able to secure a reduction in the foreign firm's output, x_{22}. This is unattainable by the

domestic firm because -- by construction of the Nash equilibrium in quantities -- the domestic firm takes x_{22} as a constant.

The result has come in for a good deal of criticism as well as attention, and it turns out not to be at all robust. Suppose that competition in the industry under study is intense, so the industry is better characterised by a Nash equilibrium in prices than in quantities. We can write profits as $\pi_{12}(p_{12}, p_{22}, t_1)$ and derive, analogous to (11),

$$\frac{dW_{12}}{dt_1} = \frac{\partial \pi_{12}}{\partial p_{22}} \frac{dp_{22}}{dt_1} + t_1 dx_{12} \tag{12}$$

But now $\partial \pi_{12}/\partial p_{22} > 0$ and $dp_{22}/dt_1 > 0$ -- an increase in rival's price raises profits, and an increase in the export tax raises rival's price. The optimal policy is therefore an export tax. Results are reversed as we change the form of the game from quantity to price competition. Whereas in Cournot competition government's power (relative to the home firm's) comes from ability to induce a reduction in rival's quantity, it now comes from an ability to induce an increase in rival's price. The former is achieved by an export subsidy and the latter by an export tax. (This result is developed in Eaton and Grossman (1986)).

The Brander and Spencer result is sensitive not only to the form of competition, but also to the number of firms, to entry and exit possibilities, and to other aspects of the model. For example, increasing the number of domestic firms in Cournot competition turns the optimal policy from an export subsidy to a tax -- as we know it must when the number of firms is large enough for the equilibrium to approach perfect competition. These and other extensions are reviewed in Brander (1995).

d; Entry and exit:

So far we have held the number of firms operating in each country constant. What happens if entry and exit is possible? Permitting the number of firms to be endogenous is

important if theory is to explain 'long run' trade patterns. It also turns out to yield very clear cut welfare and policy conclusions. (This section draws on the analysis of Venables (1985)).

We let the number of firms in each country be denoted n_1 and n_2. They all produce homogeneous output so, using the inverse demand functions we can express prices in each country as:

$$p_1 = P(\sum_{k=1}^{n_1} x_{11}^k + \sum_{k=1}^{n_2} x_{21}^k), \qquad p_2 = P(\sum_{k=1}^{n_1} x_{12}^k + \sum_{k=1}^{n_2} x_{22}^k). \tag{13}$$

where the superscript is an index labelling the firm, sums are over all firms in country 1 and in country 2 and, for simplicity, we assume identical demand functions in both countries.

Profits of a representative firm in country 1 are

$$\pi_1^k = p_1 x_{11}^k + (p_2 - t_1)x_{12}^k - c_1(x_{11}^k + x_{12}^k) - f_1 \tag{14}$$

where the technology (common to all firms in each country) has increasing returns to scale, captured by constant marginal cost c_1 and fixed cost f_1.

Given the number of firms in each country we solve for Cournot-Nash equilibrium in the usual way, by finding outputs implicitly defined by first order conditions of the form,

$$p_1 + x_{11}^k \frac{\partial P(.)}{\partial x_{11}} = c_1, \qquad p_2 + x_{12}^k \frac{\partial P(.)}{\partial x_{12}} = c_1 + t_1 \tag{15}$$

If all firms in each country are symmetric then x_{ij}^k will be the same for all k, and we can drop the superscripts. Solving for quantities gives the 'short run' equilibrium -- i.e. the equilibrium output levels and prices conditional upon the number of firms. We could use this information to express 'short run' profits as a function of numbers of firms, n_1 and n_2, but it turns out to be more insightful to make a different substitution. The short run equilibrium defines a relationship between numbers of firms and prices and, inverting this relationship, we can

express short run profits as a function of prices p_1 and p_2. We write this relationship $\Pi_1(p_1,p_2)$, $\Pi_2(p_1,p_2)$.[3]

Turning to long run equilibrium, entry and exit imply the following equations:

$$\begin{aligned} &\Pi_1(p_1, p_2) \leq 0, \quad n_1 \geq 0, \quad \textit{complementary slack}, \\ &\Pi_2(p_1, p_2) \leq 0, \quad n_2 \geq 0, \quad \textit{complementary slack}. \end{aligned} \tag{16}$$

(we assume 'large numbers' of firms so that the integer problem is ignored).

Equilibrium is illustrated on figure 2.4. The relationships $\Pi_1(p_1,p_2) = 0$ and $\Pi_2(p_1,p_2) = 0$ are illustrated. To the north-east of each of these relationships prices are high and profits are positive, and to the south-west prices are low and profits negative. Each can be thought of as mapping out what price must be in one market, given price in the other, if the representative firm from country 1 (resp. 2) is to make zero profits. Which of the curves is steeper? A decrease in p_2 must be compensated by an increase in p_1 if firms are to continue to break-even. If there are positive trade costs, t, this increase must be larger for firms in 2 than for firms in 1, because the trade costs will imply that firms in 2 are more adversely affected by the decrease in p_2 than are firms in 1. $\Pi_2(p_1,p_2) = 0$ is therefore relatively steep as illustrated.

The equilibrium on this figure is at point E, giving prices p_1 and p_2. And given knowledge of the prices we can once again use the equilibrium relationship between prices and numbers of firms to compute the numbers of firms, n_1 and n_2. (We shall not do this in detail here, but will assume there are a positive number operating in each country -- for boundary cases see Venables (1985)).

What do we learn from this analysis? First, it is possible that there is a positive number of firms in both countries, and hence intra-industry trade survives as a long run as well as a short run phenomenon. Second, market access considerations give a determinate division of the industry between locations. To understand this, suppose that costs were the same in each country, and that t = 0. In that case the $\Pi_1(p_1,p_2) = 0$ and $\Pi_2(p_1,p_2) = 0$ loci would be

identical, and we would not have a determinate division of the industry between locations. In this model it is the presence of trade costs, t > 0, that gives the two curves different slopes. Firms do relatively well in their home market, and hence price changes in their home market have a larger effect on them than do price changes in their export market, and it is this that gives the unique equilibrium at E. (We shall return to these market access considerations in section 3c below).

The analysis also gives very clear cut results on gains from trade and trade policy. In addition to giving trade prices at point E, the figure also illustrates autarky prices p_1^a and p_2^a. To see that these are autarky prices, notice that if $p_2 = 0$, firms in 1 would make no profits in market 2 and, to break even, require market 1 price p_1^a. Similarly under autarky -- no profits are made in foreign market, so p_1^a is needed in the home market to break even.

Comparison of the trade and autarky prices tells us immediately that there are gains from trade in this model. Free trade equilibrium prices are lower than autarky prices in each country, so consumer surplus is higher. There are no profits (because of free entry and exit) nor government revenue, so changes in welfare are measured by changes in consumer surplus alone. It is worth noting that these gains from trade come both from the pro-competitive effects of trade and from increasing returns to scale. Trade increases competition, reducing price, as in the oligopoly model of section 2b. With free entry and exit price equals average cost, and the reduction in average cost is brought about by expanded firm scale. It is this additional effect that guarantees the gains from trade.

What of trade policy? Consider an import tariff by country 1. This is a tax on country 2 firms, so has the effect of shifting the $\Pi_2(p_1,p_2) = 0$ locus outwards -- the tax means that firms in country 2 require higher prices if they are to break even. This shifts the equilibrium point E to the right and downwards along the unchanged $\Pi_1(p_1,p_2) = 0$ locus, reducing p_1 and raising p_2. A tariff therefore <u>reduces</u> the consumer price in the country that imposes it. The reason for this perverse result is that the tariff, as a tax on foreign firms, reduces the

equilibrium value of n_2. This in turn raises the profitability of firms in 1, and raises n_1. With positive t, and the consequent 'home market bias' in firms' sales, the net effect of these changes is to increase supply to country 1 and reduce it to country 2, with the price effects we have noted. The policy implications are immediate. Each country wants to increase its tariff to secure the reduction in price (and, as a bonus, the tariff revenue). Of course, this policy is a prisoners' dilemma for governments -- if pursued by both governments it would lead to autarky, a position we know to be Pareto inferior to free trade.

The model also yields an unambiguous result on the effects of a small export subsidy. A country 1 export subsidy benefits firms in 1, shifting the $\Pi_1(p_1,p_2) = 0$ locus inwards and reducing p_1; it can be shown that, for a small export subsidy, the benefit this brings outweighs the direct revenue cost of the subsidy.

2e: Segmented or integrated markets.

We have so far maintained the assumption of market segmentation. How is the analysis different if competition takes place in a single integrated market, i.e., each firm chooses a single value of the strategic variable for the world as a whole?

In the case of homogenous product price competition, it turns out that, if there are transport costs, then no pure strategy equilibrium exists. Mixed strategy equilibria can be found, and these support international trade with positive probability. (See Venables (1994)). Quantity competition in segmented markets is more straightforward, see Markusen (1981). The results in the entry/exit model of section 2d depend critically on the 'home market bias' of sales, and this ceases to be apply if there is a single integrated market. Horstman and Markusen (1986) and Markusen and Venables (1988) look at this case.

Which assumption is the more appropriate? Presumably some of firms' strategic variables are chosen at a world wide level -- eg R&D, model range, and perhaps capacity -- while others -- probably price -- are chosen at the level of particular markets. Venables (1990)

and Ben-Zvi and Helpman (1992) address this by looking at multi-stage games with different decisions being taken at different levels.

3. Product differentiation and monopolistic competition:

In this section we turn to the literature on international trade in industries which are monopolistically competitive, containing firms that each produce a different variety of product. The analytical structure is that of 'Spence-Dixit-Stiglitz' product differentiation in which the industry has a large number of symmetric varieties. (See Dixit and Stiglitz (1977), Spence (1976)). Firms gain market power from the fact that they are monopolists in their own variety, and researchers typically make the simplifying assumption that this is the only source of market power: firms ignore the effects of their actions on industry aggregate variables, so that there is no strategic interaction between firms. The distinctions that were so important in the preceding sections cease to be applicable.

This family of models has proved remarkably tractable, and it provides the basis of a full general equilibrium synthesis of imperfect competition and comparative advantage trade. In section 3a we set out the basic model, and we do so in a traditional two country, two industry, two factor framework. In 3b we illustrate the Helpman-Krugman (1985) synthesis of comparative advantage and intra-industry trade, in which the location of production and pattern of net trades is determined by factor endowments in a Heckscher-Ohlin fashion. In 3c we look at the effects of trade costs on the equilibrium pattern of trade. If there are trade costs then 'market access' considerations become important in determining the location of industry -- firms will want to locate in countries with large markets for their products. This provides an alternative way of determining the location of industry and pattern of trade. Section 3d uses this apparatus to study the effects of economic integration, looking both at its effects on the pattern of trade and on welfare.

The analysis of sections 3b and 3c provide alternative ways of determining the location

of industry -- based on factor endowments and market access respectively. In section 3e we develop the general theory in which both forces are present. This has interesting implications for international factor prices, and in section 3f these are analysed. It turns out that an increase in economy's endowment of a factor may raise the return to that factor, meaning that factor mobility is destabilizing. This approach provides a generalisation of Krugman's (1991a,b) work on economic geography.

3a. The model.

As in section 2, the two countries are labelled 1 and 2, and country specific variables are subscripted i, j = 1, 2. Each country is endowed with quantities L_i and K_i of two factors, the prices of which are w_i and r_i.

There are two production sectors. The z-sector is perfectly competitive and produces output which is freely tradable and will be used as numeraire. Denoting the quantity of z output in country i by z_i and its unit cost $c(w_i,r_i)$, z sector activity satisfies,

$$c(w_i, r_i) \leq 1, \qquad z_i \geq 0, \quad \textit{complementary slack}, \qquad i = 1,2. \tag{17}$$

The x-sector is imperfectly competitive, containing firms that produce differentiated products. We model product differentiation in the familiar Dixit-Stiglitz manner, so can form a CES price index for x-sector supply in each country, and we denote these price indices s_i; this price index can be thought of as a unit expenditure function for the x-sector alone. Consumer preferences between sectors are described by a homothetic expenditure function defined on the numeraire good and the x-sector price index, $e(1,s_i)u_i$, where u_i is utility. In equilibrium, income comes only from sale of factors, so the budget constraint takes the form,

$$e(1, s_i)u_i = w_iL_i + r_iK_i, \qquad i = 1, 2. \tag{18}$$

x-sector products can be supplied by national firms, the number of which in each

country is n_i (i=1,2). Country i firms set the same producer price p_i for sales in both markets, but iceberg trade costs mean that to secure delivery of one unit the consumer has to purchase τ units -- τ - 1 units melt in transit. The prices paid for delivery of a unit to home and export markets are therefore p_i and τp_i respectively. Each firm produces a single distinct variety of product, and we assume that, in each country, all firms are symmetric; we shall not introduce notation to identify individual firms. The quantities sold by a single firm in its home and export market are denoted x_{ii} and x_{ij} respectively.

Since market i is supplied by n_i home firms, and n_j foreign firms its price index for differentiated products takes the form,

$$s_i = \left[n_i p_i^{1-\epsilon} + n_j(\tau p_j)^{1-\epsilon}\right]^{1/(1-\epsilon)}, \qquad i,j = 1,2,\ i \neq j, \tag{19}$$

where ϵ is the elasticity of substitution between varieties, $\epsilon > 1$. The interpretation of this is perhaps seen most easily by looking at the quantity index dual to s_i. Denoting this y_i, it is,

$$y_i = \left[n_i x_{ii}^{(\epsilon-1)/\epsilon} + n_j x_{ij}^{(\epsilon-1)/\epsilon}\right]^{\epsilon/(\epsilon-1)}, \qquad i,j = 1,2,\ i \neq j, \tag{20}$$

Conditions for exact aggregation and the use of two-stage budgeting are satisfied, so the product $s_j y_j$ is equal to country j total expenditure on x-sector output. The quantity index can be interpreted as a sub-utility function, giving the utility derived from consumption of x-sector output. If $\epsilon = \infty$, then all exponents would be unity, saying that products are perfect substitutes (i.e. the industry produces a single homogeneous output) and the quantity index is simply the total volume consumed. A finite value of ϵ means that indifference curves between varieties are convex, so consumers choose to consume all varieties available. We shall assume that $\epsilon > 1$, so that consumers benefit from the introduction of new varieties, as can be seen by noting that the price index, s_i, is decreasing in numbers of varieties on offer.

Using the price index in the expenditure function we can find the demand for each

variety of differentiated product. Shephard's lemma gives,

$$x_{ii} = p_i^{-\epsilon} s_i^{\epsilon-1} E_i, \qquad x_{ij} = p_i^{-\epsilon} \tau^{1-\epsilon} s_i^{\epsilon-1} E_j,$$
$$E_i \equiv (w_i L_i + r_i K_i)\, s_i e_s(1,s_i)/e(1,s_i). \tag{21}$$

E_i is country i expenditure on x-sector products in aggregate, and a subscript on a function denotes partial differentiation. Although (21) gives quantities demanded, it should be noted that the quantities of the traded goods consumed, c_{ij}, is

$$c_{ij} = p_i^{-\epsilon} \tau^{-\epsilon} s_i^{\epsilon-1} E_j, \tag{22}$$

The volume produced and shipped, x_{ij}, is τ times greater than this, giving equation (21).

Technologies and profit maximisation are as follows. The profits of a single national firm in country i are,

$$p_i x_{ii} + p_i x_{ij} - b(w_i, r_i)[x_{ii} + x_{ij} + f] \tag{23}$$

where $b(w_i, r_i)$ is marginal production cost, and $b(w_i, r_i)f$ is fixed cost. Profit maximisation gives price,

$$p_i\left(1 - 1/\tilde{\epsilon}_{ij}\right) = b(w_i, r_i) \tag{24}$$

where $\tilde{\epsilon}_{ij}$ is the perceived elasticity of demand for the firms sales in market j. The form this takes depends on the nature of competition between firms. The Nash equilibrium in prices has:

$$\tilde{\epsilon}_{ij} = \epsilon(1 - m_{ij}) + \eta m_{ij} \tag{25}$$

where m_{ij} is the market share of a single firm from country i in market j, and η is the aggregate industry elasticity of demand, i.e. the elasticity of y_j with respect to s_j (we assume $\eta > \epsilon$). The

perceived elasticity is therefore a weighted average of the elasticity of demand for the firm's own variety, ϵ, and the elasticity of demand for the industry output in aggregate, η. The Nash equilibrium in quantities has:

$$\frac{1}{\tilde{\epsilon}_{ij}} = \frac{(1 - m_{ij})}{\epsilon} + \frac{m_{ij}}{\eta} \tag{26}$$

We shall frequently employ the 'large group' assumption, which says that market shares, m_{ij}, are small enough to be ignored by firms. The perceived elasticity of demand then reduces to the parameter ϵ.

Entry and exit of firms ensures non-positive profits in equilibrium. If we make the large group assumption, then price cost mark ups are the same in both markets. Using the pricing equation, (24) to eliminate p_i from the definition of profits we can therefore express the industry equilibrium condition as,

$$f(\epsilon - 1) \geq x_{ii} + x_{ij}, \quad n_i \geq 0, \quad \textit{complementary slack.} \tag{27}$$

Total sales must reach level $f(\epsilon - 1)$, where the values of x_{ii} and x_{ij} are determined by demand equations (21). The power of the large group assumption is immediately apparent. With this assumption $\tilde{\epsilon}$ is simply equal to the parameter ϵ, so equilibrium firm scale is a constant, depending only on parameters of the model.

To complete characterization of equilibrium we need only specify factor market clearing. This takes the form,

$$L_i = z_i c_w(w_i, r_i) + n_i\left[x_{ii} + x_{ij} + f\right] b_w(w_i, r_i) \tag{28}$$

$$K_i = z_i c_r(w_i, r_i) + n_i\left[x_{ii} + x_{ij} + f\right] b_r(w_i, r_i) \tag{29}$$

The terms on the right hand side give factor demands associated with z-sector and with x-sector production.

3b: Factor endowments and the integrated equilibrium

If there is completely free trade ($\tau = 1$) and we make the large group assumption, then the model is that of Helpman and Krugman (1985). We can use the 'integrated equilibrium' approach to illustrate the equilibrium. Suppose that the world in not divided into two economies, each with its own endowment, but is instead a single economy. We can find the equilibrium of this economy, and we shall call it the 'integrated equilibrium'. We find, amongst other things, the techniques of production in use in each sector, and these labour capital ratios will be denoted λ_z and λ_x respectively. We shall label factors such that the x-sector is capital intensive.

Now consider figure 3.1. The dimensions of the box are the world endowment, with capital on the horizontal and labour on the vertical. We divide this between the two countries, with origins for country 1 in the southwest corner and country 2 in the north east. A point in the box therefore describes a division of the world endowment between the two countries.

We now pose the following question. Suppose the world is divided into two countries, and in particular, the endowment is divided between these countries. If factors are assumed non-traded and goods are freely traded, can trade reproduce the integrated equilibrium? The answer is yes, providing the endowments lie in the parallelogram formed by O_1BO_2D. The sides of this parallelogram are rays from the origins (both O_1 and O_2) with gradient equal to labour capital ratios; operating the x industry therefore employs capital and labour in amounts given by the line O_1x, and operating the z industry employs capital and labour in amounts given by the line O_1z. If the endowment is in the parallelogram then there are non-negative output levels for each industry in each country which fully employ both factors in both

countries. This means that the free trade equilibrium reproduces the integrated equilibrium -- we have full employment of both factors at the same techniques of production and factor prices, this generating the same costs and prices, income and demand, and supply of all goods.

The parallelogram is referred to as the diversification set, i.e. the set of endowments in which both industries are active in both countries. In the case we are studying, with $\tau = 1$, trade can reproduce the integrated equilibrium if and only if the factor endowment lies inside this set. If trade reproduces the integrated equilibrium then factor prices in the two countries are the same, so the diversification set is also the factor price equalisation set.

Outside the diversification set the configuration is similar to that described by Dixit and Norman (1980). If economy 1 is very labour abundant then $n_1 = 0$, and if 2 is very capital abundant, then $z_2 = 0$. The dividing line between these regions is line AB. North-east of this line economy 2 is not large enough to produce world demand for x-sector output, so we have $n_1, n_2 > 0$; south-west of the line economy 1 is not large enough to produce world demand for z, so we have $z_1, z_2 > 0$.

If we restrict attention to the FPE set then the pattern of trade can be found easily. Suppose the endowment is at E. The factor content of consumption will then be at corresponding point C, constructed to satisfy two properties. First, C must lie on the main diagonal. This is because countries have identical homothetic preferences, so they consume x- and z- sector output in the same proportions; they must also therefore consume the factors embodied in this consumption in the same proportions. Second, the position of C relative to E is determined by the fact that C lies on the line through E with gradient r/w. This is the budget constraint, saying that the value of the endowment equals the value of factors embodied in consumption.

The vector EC is the difference between a country's factor endowment and factor content of consumption, and is therefore the factor content of its trade. As drawn, country 1 is a net importer of capital and exporter of labour, embodied in traded goods. This factor

content of trade is achieved by 1 exporting the labour intensive good, z, and importing capital intensive x-sector products. The pattern of inter-industry trade is therefore exactly as in the Heckscher-Ohlin model.

While inter-industry trade is as predicted in the Heckscher-Ohlin model, this model also supports intra-industry trade. x-sector firms are operating in each country, and each firm sells its output in each country, at levels determined by the demand functions, equations (21). The volume of this trade is largest if endowments in both countries are identical (the centre of the endowment box). At this point there is no inter-industry trade (because relative endowments are the same), and the x-sector industry is divided equally between the countries. Each firm sells half its output in each country, meaning that half of x-sector output is traded internationally.

c: Trade costs and market access.

The Helpman and Krugman analysis is an elegant synthesis of factor endowment and intra-industry trade, but the assumption of completely free trade means that each importing firm has exactly the same market share as each domestic firm, implying that the volume of trade may be as large as half the volume of output in a two-country model, or fraction (N-1)/N the volume of output if there are N identical countries. This is an order of magnitude larger than trade flows actually observed. How are things changed if we allow for a 'home market bias' in sales, by introducing positive trade or transport costs, $\tau > 1$?

The equilibrium location of the industries must always satisfy two sorts of conditions. The first are factor market conditions, stating that the industries must generate factor demands sufficient to employ the factor supplies in each economy. The second are product market conditions, stating that supply of output must equal demand in each economy.

If trade is completely free, as in the preceding section, then the product market conditions are relaxed. World supply must equal world demand for each product, but -- from

the point of view of the product market -- the location of production is immaterial, as output can be transported costlessly to supply either market. This is the basis of traditional models of trade, as well as Helpman and Krugman. Providing trade is completely costless, the division of industries between countries is determined by factor market, not product market considerations.

If there are trade barriers, then we have to pay attention to supply and demand in each market. Expenditure levels in each market then also have a role in determining the location of production -- other things being equal, countries with a large expenditure on the x-sector industry will have a large volume of production in this industry. To see this it is helpful to derive an explicit equation for the value of x-sector production in each location. We can do this as follows. Using the demand equations (21) in the zero profit condition (27) we obtain;

$$f(\epsilon - 1) = p_i^{-\epsilon} s_i^{\epsilon-1} E_i + \tau^{1-\epsilon} p_i^{-\epsilon} s_j^{\epsilon-1} E_j, \qquad i,j=1,2,\ i \neq j. \tag{30}$$

We can solve these two equations to express the terms $s_i^{\epsilon-1} E_i$ in the following form:

$$s_i^{\epsilon-1} E_i = \frac{f(\epsilon - 1)\left[p_i^{\epsilon} - p_j^{\epsilon}\tau^{1-\epsilon}\right]}{1 - \tau^{2(1-\epsilon)}}, \qquad i,j = 1,2, \quad i \neq j. \tag{31}$$

Suppose that factor prices are the same in the two countries, so $p_1 = p_2$, and we denote the common value $\bar{p}$. The x-sector price index (equation (19)) can therefore be expressed as,

$$s_i^{1-\epsilon} \bar{p}^{\epsilon} = n_i \bar{p} + n_j \bar{p} \tau^{1-\epsilon}. \tag{32}$$

From (31) and (32) we obtain

$$n_1 \bar{p} f(\epsilon - 1) = \frac{E_1 - \tau^{1-\epsilon} E_2}{1 - \tau^{1-\epsilon}}, \qquad n_2 \bar{p} f(\epsilon - 1) = \frac{E_2 - \tau^{1-\epsilon} E_1}{1 - \tau^{1-\epsilon}}. \tag{33}$$

The left hand side of these equations is the value of x-sector production in countries 1 and 2

(recall that the zero profit level of output is attained at scale $f(\epsilon-1)$). The equations tell us how the value of production depends trade cost and expenditure levels in each country.

What do we learn from these equations? First, if there is completely free trade, $\tau = 1$, then the equations are not defined. As we have already seen, in this case the location of production is determined entirely by factor supply, and product market access considerations are irrelevant.

Second, if $\tau > 1$ (and hence $1 > \tau^{1-\epsilon}$), then production is skewed towards the location with the larger market. This is unsurprising, but a stronger result also holds. Differences in expenditures lead to a proportionately larger difference in output levels. To see this, notice that production in country 1 relative to country 2 can be expressed (taking the ratio of equations (33)) as

$$\frac{n_1}{n_2} = \frac{E_1/E_2 - \tau^{1-\epsilon}}{1 - \tau^{1-\epsilon} E_1/E_2}, \tag{34}$$

This relationship is illustrated in figure 3.2. There is production in both countries only if $E_1/E_2 \in (\tau^{1-\epsilon}, \tau^{\epsilon-1})$. The relationship intersects the 45° line from below, illustrating that changes in E_1/E_2 are associated with proportionately larger changes in n_1/n_2. Furthermore, the *smaller* is τ (larger is $\tau^{1-\epsilon}$) the steeper is the function indicating increasing sensitivity of the location of production to expenditure differences.

This establishes that market size is a determinant of the location of production and pattern of trade. Countries with large markets in a particular sector will have a disproportionately high share of that sector's production, and will be a net exporter of that product. The reason is market access -- firms will want to locate in the large market. The intuition can also be seen by considering what would happen if the relative location of production was the same as the relative location of demand ($n_1/n_2 = E_1/E_2$). The large economy would then have a lower import penetration ratio than the small, so entry of firms

would be profitable in the large rather than in the small.

3d: Regional economic integration;

In sections 3e and 3f we shall synthesise factor market and market access theories of the location of industry. In this section we digress to look at an application of this model. What does it tell us about the effects of regional economic integration schemes? We look first at the positive effects of integration and show how it may cause 'production shifting' between economies. We then turn to the welfare effects of integration. Fuller analysis of this topic is provided in Baldwin and Venables (1995).

The positive effects of integration on the location of industry can be derived from the analysis of section 3c. Suppose there are only two economies, and trade barriers between these countries are reduced. As τ falls, which country gains industry? It can readily be shown by differentiation of (34) that, if we take expenditure levels E_i as exogenous, then reductions in τ shift industry towards the economy with the larger expenditure (the n_1/ n_2 curve of figure 3.2 rotates anti-clockwise around the point (1,1). This is of course just an application of the result that large economies will be net exporters of the product of the imperfectly competitive industry. The result suggests that a process of economic integration may tend to draw activity into countries or regions with good market access, at the expense of peripheral regions with smaller local markets.

What of a preferential regional integration scheme? To analyse this we need a three country model, in which two countries, the union members, have lower trade barriers with each other than they do with the third country. Suppose that the two union members have the same expenditure level, E_1, and that the third country has expenditure E_3. Iceberg trade costs between union members are τ, and with the third country are θ. Proceeding analogously to the derivation of equations (30) to (33) we can derive output levels in one of the union countries and in the third country as:

$$n_1 \bar{p} f(\epsilon - 1) = \frac{E_1}{1 - \theta^{1-\epsilon}} - \frac{E_3 \theta^{1-\epsilon}}{1 + \tau^{1-\epsilon} - 2\theta^{1-\epsilon}},$$

$$n_3 \bar{p} f(\epsilon - 1) = \frac{E_3 (1 + \tau^{1-\epsilon})}{1 + \tau^{1-\epsilon} - 2\theta^{1-\epsilon}} - \frac{2E_1 \theta^{1-\epsilon}}{1 - \theta^{1-\epsilon}} \tag{35}$$

Differentiating these equations it can be shown that integration between 1 and 2 -- a reduction in τ (increase in $\tau^{1-\epsilon}$) holding θ constant -- raises n_1 $(=n_2)$ and reduces n_3. In other words, integration causes 'production shifting' into the union from third countries. It works because integration raises the profits of firms in the union countries, attracting entry. But these entrants also supply country 3, reducing profits of firms located in country 3, causing their exit which in turn attracts further entry in 1 and 2, and so on.

What of the welfare effects of integration? If we continue to assume iceberg trade costs and make the large group assumption, then the welfare effects can be derived directly. For example, suppose that there are just two integrating economies, and that they are symmetric, so $p \equiv p_1 = p_2$. Then the equilibrium value of the price index, as given by equation (31), becomes

$$s_i^{\epsilon-1} E_i = \frac{f(\epsilon - 1)p}{1 + \tau^{(1-\epsilon)}}, \qquad i,j = 1,2, \qquad i \neq j. \tag{36}$$

A reduction in iceberg trade costs by amount $d\tau$ causes change in the price index,

$$\frac{ds_i}{s_i} = m\frac{d\tau}{\tau}, \qquad m \equiv \frac{\tau^{1-\epsilon}}{1 + \tau^{1-\epsilon}} \tag{37}$$

where m is the share of imports. $-ds_i/s_i$ is a direct measure of the welfare change -- it gives the proportional reduction in the price index in the sector affected by the integration. Equation (37) says that this welfare gain is just equal to the reduction in trade costs times the proportion

of output directly affected by this reduction. In other words, the only gain (to a first order approximation) is the direct cost saving from the reduction in the real trade cost. Firm scale does not change (it is fixed by the large group assumption) and the number of firms operating in the industry does not change (as can be seen by totally differentiating (32) and using (37)).

This result is rather disappointing -- the absence of any induced welfare changes in a model of imperfect competition is rather surprising. However, the result turns on two very strong assumptions that have been employed -- iceberg trade costs and the large group assumption.

Suppose that trade barriers take the form of tariffs, rather than real trade costs. The effects in this case can be derived by using demand function (22) instead of (21) and we derive, analogous to (37),

$$\frac{ds_i}{s_i} = m\frac{\epsilon}{\epsilon - 1}\frac{d\tau}{\tau}, \qquad m \equiv \frac{\tau^{-\epsilon}}{1 + \tau^{-\epsilon}} \tag{38}$$

where the import share, m, no longer includes the iceberg element. The price index change is larger in this case, essentially because the induced increase in trade volume no longer uses up real resources. To calculate the welfare change we must also compute the change in government revenue. If this is done, it can be shown that there is a net welfare gain from the integration.

The 'large group' assumption serves to rule out any 'pro-competitive' effects of economic integration. As we have seen, this assumption says that the perceived elasticity of demand is equal to the constant parameter, ϵ. As well as ruling out 'pro-competitive' effects, this assumption also means that firm size is constant, and unchanged by changes in τ, so that integration can bring no gains from fuller exploitation of returns to scale.

To relax the large group assumption we return to the pricing equations (24) - (26). Integration changes market shares, m_{ij}, this taking the form of an increase in each firm's share

in export markets and a fall in its home market share. Under either quantity or price competition this raises the perceived elasticity of demand in the home market and reduces it in export markets, causing a reduction in price cost margins in the home market, and increase in theses margins on export sales. How does this affect profits? If home sales are larger than sales in each export market, then the net effect is a reduction in profits -- the reduction in margins occurs on high volume and the increase on small. (For fuller analysis of this see Baldwin and Venables (1995)). The reduction in profits causes exit of firms, and some loss of product variety. However, it also expands the scale of remaining firms, and reduces their average costs. Integration therefore both increases the intensity of competition, as firms invade each others markets, and increases firm scale and reduces average costs, as there is a change in the number of firms in response to the change in profit levels.

Attempts to quantify these effects are given in Smith and Venables (1988). This paper also investigates the implications of integration being associated with a change in the nature of competition, switching the equilibrium from one with segmented markets to one with integrated markets. If this were to occur then the market share terms in the pricing equations would no longer be shares of each firm in each market, but instead shares of each firm in the entire market of the union. This amplifies the effects we have just outlined. Firms lose relatively protected positions in their domestic market, as competition takes place in the union market as a whole.

3e; Combining factor endowments and market access;

We have seen in sections 3b and 3c that it is possible to determine the location of industry in two different ways -- through factor endowments, in a Heckscher-Ohlin manner or, if there are positive trade costs, through market access considerations. What happens when we put these two approaches together?

The easiest way to address this question is to ask what happens to the FPE set when

we add trade costs to the model of section 3a. Let us call FPE factor prices $w_1 = w_2 = \bar{w}$, and $r_1 = r_2 = \bar{r}$ giving x-sector goods price $\bar{p}$. We have already seen that the value of x-sector output in each country, as determined by market access considerations, is (equation (33)):

$$n_1 \bar{p} f(\epsilon - 1) = \frac{E_1 - \tau^{1-\epsilon} E_2}{1 - \tau^{1-\epsilon}}, \qquad n_2 \bar{p} f(\epsilon - 1) = \frac{E_2 - \tau^{1-\epsilon} E_1}{1 - \tau^{1-\epsilon}}. \tag{39}$$

Note now that, in general equilibrium, expenditure E_i, depends on factor endowments, since $E_i = \sigma(\bar{w} L_i + \bar{r} K_i)$, where σ is the share of the x-sector in consumption, $\sigma \equiv s_i e_s(1, s_i)/e(1, s_i)$. We shall assume Cobb-Douglas preferences, so that σ is a constant.

This means that in the FPE set expenditure is a linear function of endowments, implying that x-sector production is also a linear function of endowments. However, this dependence of production on factor endowments is via expenditures, not through the factor market.

What about factor market considerations? Assuming that both the z- and x-sectors are operating in both countries and denoting technical coefficients at FPE factor prices by $\bar{b}_w$ etc, we can use factor market clearing conditions, (28) and (29), to derive,

$$n_1 \bar{p} f(\epsilon - 1) = \bar{b} \left(\frac{\bar{c}_w K_1 - \bar{c}_r L_1}{\bar{c}_w \bar{b}_r - \bar{b}_w \bar{c}_r} \right), \qquad n_2 \bar{p} f(\epsilon - 1) = \bar{b} \left(\frac{\bar{c}_w K_2 - \bar{c}_r L_2}{\bar{c}_w \bar{b}_r - \bar{b}_w \bar{c}_r} \right). \tag{40}$$

This gives a standard Heckscher-Ohlin relationship, and says that the value of output in each country depends on factor endowments in each country and on technical coefficients.

Equations (39) and (40) give the product and factor market determinants of x-sector production. Requiring them both to be satisfied restricts endowments, and this restriction defines the FPE set. Eliminating n_1 (or n_2 -- the two pairs of equations are not independent) from these equations, noting that both expressions are linear in K_i and L_i and using world endowments $K = K_1 + K_2$ and $L = L_1 + L_2$ we can characterise the FPE set as a linear

relationship between K_1 and L_1. The FPE set is no longer the full diversification set of figure 3.1, but instead a one-dimensional subspace of this set.[4]

The expression for the FPE set is straightforward, and we need only note that it includes the point at which the two economies are identical (the midpoint of the main diagonal of figure 3.1), is linear, and has gradient,

$$\frac{dL_1}{dK_1} = \frac{\bar{c}_w\bar{b}(1-\tau^{1-\epsilon}) - \sigma\bar{r}\Delta(1+\tau^{1-\epsilon})}{\bar{c}_r\bar{b}(1-\tau^{1-\epsilon}) + \sigma\bar{w}\Delta(1+\tau^{1-\epsilon})} = \frac{\bar{c}_w\bar{b}(1-\sigma) + \bar{b}_w\bar{c}\sigma - \tau^{1-\epsilon}(\bar{c}_w\bar{b}+\sigma r\Delta)}{\bar{c}_r\bar{b}(1-\sigma) + \bar{b}_r\bar{c}\sigma - \tau^{1-\epsilon}(\bar{c}_r\bar{b}-\sigma w\Delta)} \tag{41}$$

where Δ is the determinant of the matrix of technical coefficients, $\Delta \equiv \bar{c}_w\bar{b}_r - \bar{b}_w\bar{c}_r > 0$, positive since the x- sector is capital intensive. (The two different forms given for the gradient are useful in different contexts, and can be derived from each other using homogeneity of the cost functions).

This is illustrated in Fig 3.3 in which the line segment *bc* is the part of the FPE set in the diversification set (this figure is constructed with the same parameters as fig. 3.1 except that τ= 1.2). It goes through the midpoint of the endowment box and has gradient given by (41). The gradient depends on the level of transport costs in the following way. As $t \to \infty$, so the gradient of *bc* goes to that of the O_1O_2 diagonal of the box, because under autarky there is FPE only if the endowment ratios of the two economies are equal. (This can be seen by letting $\tau \to \infty$ in the second part of equation (41) which then gives the ratio of world demand for labor to world demand for capital).

With finite τ, the gradient of *bc* is less than that of the O_1O_2 diagonal, as illustrated in Fig 3.3. The intuition for this can be seen by considering a point below the O_1O_2 diagonal. At such a point country 1 is relatively well endowed with capital, and hence has relatively much x-sector production. If there are transport costs, this is consistent with FPE only if country 1 has relatively high demand for x products, i.e., the point is closer to O_2 than to O_1. *bc* must therefore cut the O_1O_2 diagonal from above, as illustrated.

This market access affect is powerful when transport costs are large, but weak as they become small. Reducing τ therefore reduces the gradient of bc, rotating the FPE set clockwise around the central point, and at some value of τ the FPE set is horizontal (see the first part of equation (41)). At still lower values of τ the FPE set has negative gradient, and as we go to free trade ($\tau \to 1$) the gradient of the FPE set tends to the ratio of factor prices, $-\bar{r}/\bar{w}$; factor price equalisation then occurs when the two economies have the same income levels and hence the same consumer demand for x-sector products.

Our theoretical discussion has hinged on diversified production, with both sectors active in both countries. Figure 3.3 also traces out the continuation of the FPE set outside the region in which production is diversified. The entire FPE set is made up of the three segments, O_1b, bc, and cO_2.

What do we know about equilibrium at endowments outside the FPE set? At endowments above the FPE set we have $p_1 > p_2$. This is because country 1 is relatively poorly endowed with capital, the factor used intensively in x-sector production, and, relative to this, has a large total endowment and hence large demand for x-sector output. The x-sector price differences illustrated in the figure are associated with factor price differences. Since the x-sector is relatively capital intensive, $r_1 > r_2$ and $w_1 < w_2$ above the FPE line.

What about the pattern of trade? At the central point of the endowment box there is no net trade, although there is intra-industry trade, this amounting to less than half of x-sector production because of the presence of trade costs. If we move along the FPE set to the north-east of the central point what happens? Country 1 gets larger and more capital intensive, both of these considerations making country 1 a net exporter of x-sector output.

Now consider moving north-east from the central point of the endowment box not along the FPE set, but instead along the diagonal, towards O_2. Countries have the same relative factor endowments, but country 1 becomes the larger country. Country 1 x-sector production therefore expands and country 2 contracts because of market access

considerations. We therefore observe country 1 being a net exporter of x-sector output. This has the interesting implication that each country is a net exporter of the product which is relatively intensive in its expensive factor. Essentially, market access considerations are more powerful than factor price differentials in determining location.

Moving along the O_1O_2 diagonal the two economies have the same relative factor intensities, so market access considerations determine the pattern of trade. What happens as we move away from the diagonal? As we move far enough away so factor market considerations come to dominate, and the pattern of net trades reverts to that expected from Heckscher-Ohlin theory. For example, at a point near the top of the fig. 3.2 and to the right of centre, country 1 is large, but very labour abundant relative to country 2; it will then export z, the labour intensive good.

This is all illustrated in figure 3.2 . On the curved line there is intra-, but not inter-industry trade. Below it country 1 is a next exporter of x-sector output, and above it a net importer -- by virtue of either small size or relative labour abundance.

3f. Factor prices, factor mobility, and 'new economic geography'.

We have so far assumed that neither capital nor labour are internationally mobile. What happens if these factors can move in response to international factor price differences? We have already seen how x-sector production is drawn towards large economies for market access reasons. But the locational attractiveness of large economies raises factor demands and factor prices in these economies, so attracting factor inflow and making large economies even larger. This suggests a 'positive feedback' or 'cumulative causation' mechanism, under which factor mobility may be destabilising, causing economic activity to agglomerate in one country. These forces were studied in Krugman (1991a,b). In this section we generalise Krugman's analysis by placing it within the 2x2x2 framework we have developed.

So far, we have looked at factor prices in terms of the numeraire. However, price

indices s_i differ between countries, and so therefore does the cost of living, as measured by the unit expenditure functions, $e(1,s_i)$. We shall look at cases where mobile factors consume in the host economy, so incentives to move are determined by factor returns deflated by these expenditure functions.

Figure 3.5 makes the required modifications to figure 3.3. The dashed line O_1bcO_2 in figure 3.5 gives the FPE set constructed exactly as in figure 3.3. Above this line country 1 is poorly endowed with capital relative to labour, so $r_1 > r_2$ and $w_1 < w_2$, and conversely below the line. Adjusting for differences in price indices, s_i, real factor incomes are $R_i = r_i/e(1,s_i)$ and $W_i = w_i/e(1,s_i)$, and the lines $R_1 = R_2$ and $W_1 = W_2$ plot the endowment points along which real returns to each factor are equalised. The intuition behind the positions of these lines is straightforward. Consider point *d*. At *d* factor prices are equalised in terms of the numeraire, but economy 1 has more x-sector production than economy 2 (it has a larger market and a more capital intensive endowment). We consequently have $s_1 < s_2$, so country 1 has a higher real return to both factors. This means that *d* must lie below the iso-real-w line, and above the iso-real-r line.

The arrows marked on the figure indicate the directions of factor mobility when factors move in response to real factor price differences. Labour mobility moves the endowment vertically to the iso-real-w line, $W_1 = W_2$. Capital mobility causes horizontal movements, and if the segment BC of the iso-real-r line has negative gradient as illustrated, then capital mobility moves the endowment to segments O_1B or O_2C of the iso-real-r line. The line segment *BC* is unstable with respect to capital mobility. The reason is that a move to the right of this segment causes a large increase in x-sector production in country 1 (a Rybczynski effect, as country 1 gains more of the factor intensive in the x-sector), this reducing s_1 and thereby raising the real income of factors in country 1, so attracting further capital inflow. This price index effect means that mobility of the factor intensive in the x-sector may cause divergence of

the structure of the two economies.

The comparison of labour and capital mobility indicates that divergence is associated with mobility of capital, the factor used intensively in the imperfectly competitive sector. What if both capital and labour are mobile? The only point in the figure at which real factor price equalisation occurs at point E. This point is unstable -- a small perturbation around the point leads to divergence. Full factor mobility therefore leads to one of the origins, with all activity agglomerated in one of the locations.

The instability outlined above hinges on the line segment BC having negative gradient, and this will be so if transport costs, τ, are not too large. To see this it is worth developing the special case of this model originally put forward by Krugman (1991a). Krugman specialises the model by assuming that each sector uses a single factor, so technology coefficients take the form:

$$b_w = 0, \quad b_r = 1, \quad c_w = 1, \quad c_r = 0. \tag{42}$$

The x-sector only uses capital, and the z-sector only uses labour. (Krugman refers to the x-sector as manufacturing, using manufacturing labour, and the z-sector as agriculture using peasants who are distinct from manufacturing labourers. We retain our original, if less colourful, terminology). Since the z-sector is numeraire, this assumption immediately fixes wages,

$$w_1 = w_2 = 1. \tag{43}$$

With this structure it is easy to see that the factor market condition (equation (40), using (42) and pricing equation (24)) reduces to

$$n_1 f\epsilon = K_1, \quad n_2 f\epsilon = K_2. \tag{44}$$

Krugman assumes that labour is immobile and capital is perfectly mobile in response to

differences in host country real returns[5]. This means that equilibrium returns satisfy

$$r_1/s_1^{\sigma} = r_2/s_2^{\sigma}. \tag{45}$$

where σ is the share of the x-sector in consumption, so the denominators perform the role of deflating by the local cost of living index.

We shall restrict analysis to the following question. Suppose that the x-sector is concentrated in country 1 ($n_1 > 0$, $n_2 = 0$). For what parameter values is this an equilibrium? The assumption that $n_2 = 0$ simplifies analysis greatly. First, it means that the price indices (equations (19)) collapse down to

$$s_1 = n_1^{1/(1-\epsilon)} p_1, \qquad s_2 = n_2^{1/(1-\epsilon)} \tau p_1 \tag{46}$$

Second, it means that expenditure levels in each country reduce to

$$E_1 = \sigma(r_1 K + L_1), \qquad E_2 = \sigma L_2 \tag{47}$$

where K is the world stock of capital and because country 2 has no x-sector production, it also has no capital. Furthermore, we shall assume that each country has half the world's endowment of labour, and write $L_1 = L_2 = L/2$. Since capital income equals x-sector output it takes fraction σ of world income, and labour income equals z-sector output it takes fraction $1 - \sigma$ of income. The ratio of capital income to labour income must therefore equal the ratio of expenditure shares of each good, so $r_1 K/L = \sigma/(1-\sigma)$. This means that expenditures can be rewritten as,

$$E_1 = \frac{\sigma L}{2}\left(\frac{1+\sigma}{1-\sigma}\right), \qquad E_2 = \frac{\sigma L}{2} \tag{48}$$

Third, the assumption that $n_2 = 0$ allows us to write the country 1 product market condition (equation (30), using (46)) immediately as,

$$n_1 p_1 f(\epsilon - 1) = E_1 + E_2 = r_1 K + L = L/(1 - \sigma) \tag{49}$$

We must now check whether our assumed pattern of production, $n_1 > 0$, $n_2 = 0$ is an equilibrium. It will be so if a potential entrant in country 1 would fail to make the level of sales required for break even. The expression for break even is given above as equation (30), so the condition is

$$f(\epsilon - 1) \leq p_2^{-\epsilon} s_2^{\epsilon - 1} E_2 + \tau^{1-\epsilon} p_2^{-\epsilon} s_1^{\epsilon - 1} E_1. \tag{50}$$

where the left hand side gives the scale the potential entrant must reach to break even, and the right hand side gives demand for its output. Using equations (45) - (49) this simplifies to the condition

$$1 \geq \tau^{-\sigma\epsilon}\left[\tau^{1-\epsilon}\left(\frac{1+\sigma}{2}\right) + \tau^{\epsilon-1}\left(\frac{1-\sigma}{2}\right)\right] \tag{51}$$

Providing parameters satisfy this condition, agglomeration of activity is an equilibrium. The relationship is illustrated, as a function of τ, on figure 3.6. At very high τ agglomeration is not an equilibrium.[6] This is because there is demand for x-sector output from the immobile factor, labour, in both countries, and this has to be met by local production. At free trade, $\tau = 1$, there is nominal and real FPE over the entire diversification set. However, there is a range of values of τ at which the right hand side of (51) is less than unity, and at these points agglomeration of activity in one location is an equilibrium.

This result is very extreme, and it is easy to think of considerations that would slow down or halt agglomeration forces -- adjustment costs, or the existence of other immobile factors (land). Nevertheless, the mechanism is important, and suggests that factor mobility may have perverse effects, tending to make the distribution of economic activity unequal between locations. This same mechanism has been used in the regional economics literature to

explain the development of urban centres (for example Fujita (1988)). For discussion of alternative forces with give rise to agglomeration see Krugman (1991b).

4. Concluding comments:

These notes discuss some of the main results and models from the theory of international trade under imperfect competition. They are necessarily both selective and superficial. Multinationals are conspicuous by their absence, and the reader is referred to Markusen (1995) for a recent survey. Up to date and exhaustive treatments of other topics are provided in volume III of the Handbook of International Economics, published in 1995. (See references to Brander (1995), Baldwin and Venables (1995)).

Appendix;

Simulations: Figures 3.1-3.5 were computed from simulations use the following functional forms:

$$c(w_i,r_i) = w_i^{2/3} r_i^{1/3}, \quad b(w_i,r_i) = w_i^{1/3} r_i^{2/3}, \quad e(1,s_i) = s_i^{1/2}. \tag{52}$$

World endowments are set at K = L = 1. The elasticity of demand for a single variety of differentiated product, $\epsilon = 5$. National firms have fixed input requirement f = 0.25. Values of τ are reported in the text.

Endnotes:

1. Some aspects of this analysis are developed in Dixit (1984). Another valuable reference for the policy material is Brander (1995).

2. We index tax instruments by the firm on which they bear, not the government to which revenue accrues

3. This function Π embodies equilibrium behaviour, and is different from the profit functions of section 2c

4. The Cobb-Douglas assumption is necessary for the FPE set to be linear, but not for it be a one dimensional subset of the endowment space.

5. This means that income of the mobile factor is consumed in the host not the source country -- plausible if the mobile factor is some type of labour, but less plausible if it is capital.

6. This is consistent with our discussion of the way in which changes in τ rotate the nominal FPE line (eg eqn (41)) and hence determine the slope of the real FPE line of figure 3.5

References:

R. Baldwin and A.J. Venables, (1995), 'Regional economic integration', in *Handbook of International Economics, vol III*, ed G. Grossman and K. Rogoff, Elsevier, Amsterdam.

S. Ben-Zvi and e. Helpman, 'Oligopoly in segmented markets', in G. Grossman (ed) *Imperfect Competition and international trade'* MIT, Cambridge MA.

J. Brander, (1995), 'Strategic trade policy', in *Handbook of International Economics, vol III*, ed G. Grossman and K. Rogoff, Elsevier, Amsterdam.

J. Brander and P. Krugman (1983), 'A reciprocal dumping model of international trade', *Journal of International Economics*, 15:313-323.

J. Brander and B. Spencer (1985), 'Export subsidies and market share rivalry', *Journal of International Economics*, 18:83-100.

A. K. Dixit, (1984)'International trade policy for oligopolistic industries' *Economic Journal*, 94, 1-16.

A. K. Dixit and V. Norman (1980), *The Theory of International Trade*, Cambridge, CUP.

A. K. Dixit and J.E. Stiglitz (1977), 'Monopolistic competition and optimum product diversity', *American Economic Review*, 67: 297-308.

J. Eaton and G. Grossman, (1986), 'Optimal trade and industrial policy under oligopoly', *Quarterly Journal of Economics*, 101:383-406.

M. Fujita, (1988), 'A monopolistic model of spatial agglomeration', *Regional Science and urban economics*, 12, 161-196.

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

I. Horstman and J. Markusen, (1986),'Up the average cost curve; inefficient entry and the new protectionism', *Journal of International Economics*,, 20, 225-247.

E. Helpman and P.R. Krugman, (198?) *Policy*, Cambridge MA, MIT press.

P. Krugman (1979), 'Increasing returns, monopolistic competition and international trade', *Journal of International Economics*, 9: 469-479.

(1980), 'Scale economies, product differentiation and the pattern of trade', *American Economic Review*, 70: 950-959.

(1981), 'Intra-industry specialisation and the gains from trade', *Journal of Political Economy*, 89: 959-973.

(1991a), Increasing returns and economic geography', *Quarterly Journal of Economics*

(1991b), *Geography and Trade*, MIT, Cambridge MA

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

J.R. Markusen and A.J. Venables, (1988) 'Trade policy with increasing returns and imperfect competition; contradictory results from competing assumptions, *Journal of International Economics*, 24: 299-316.

A. Smith and A.J. Venables, (1988), 'Completing the internal market in the European Community; some industry simulations', *European Economic Review*, 32, 1501-1525.

Spence, A.M., (1976),'Product selection, fixed costs and monopolistic competition' *Review of Economic Studies*, 43, 217-235.

A.J. Venables (1985), 'Trade and trade policy with imperfect competition; the case of identical products and free entry', *Journal of International Economics*, 19, 1-20.

A.J. Venables (1990), 'International capacity choice and national market games', *Journal of International Economics*, 29, 23-42.

A.J. Venables (1994), 'Tariffs and subsidies with price competition and integrated markets', *Oxford Economic Papers*, 46, 30-44.

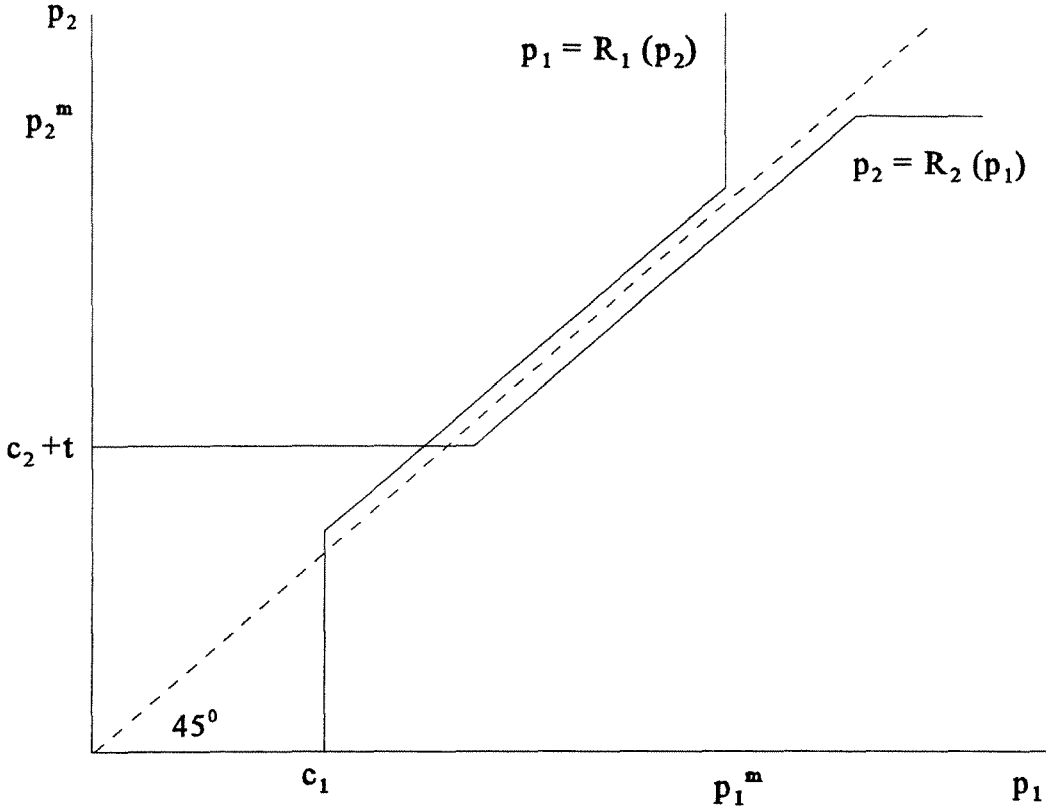

Figure 2.1

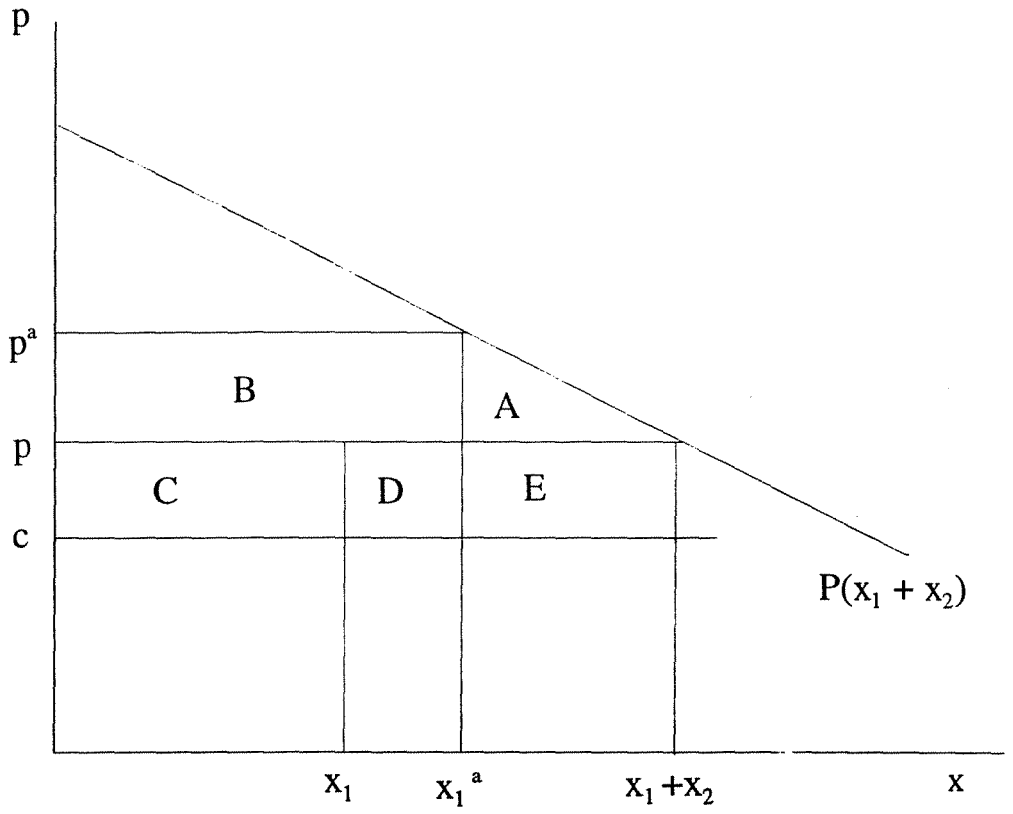

Figure 2.2

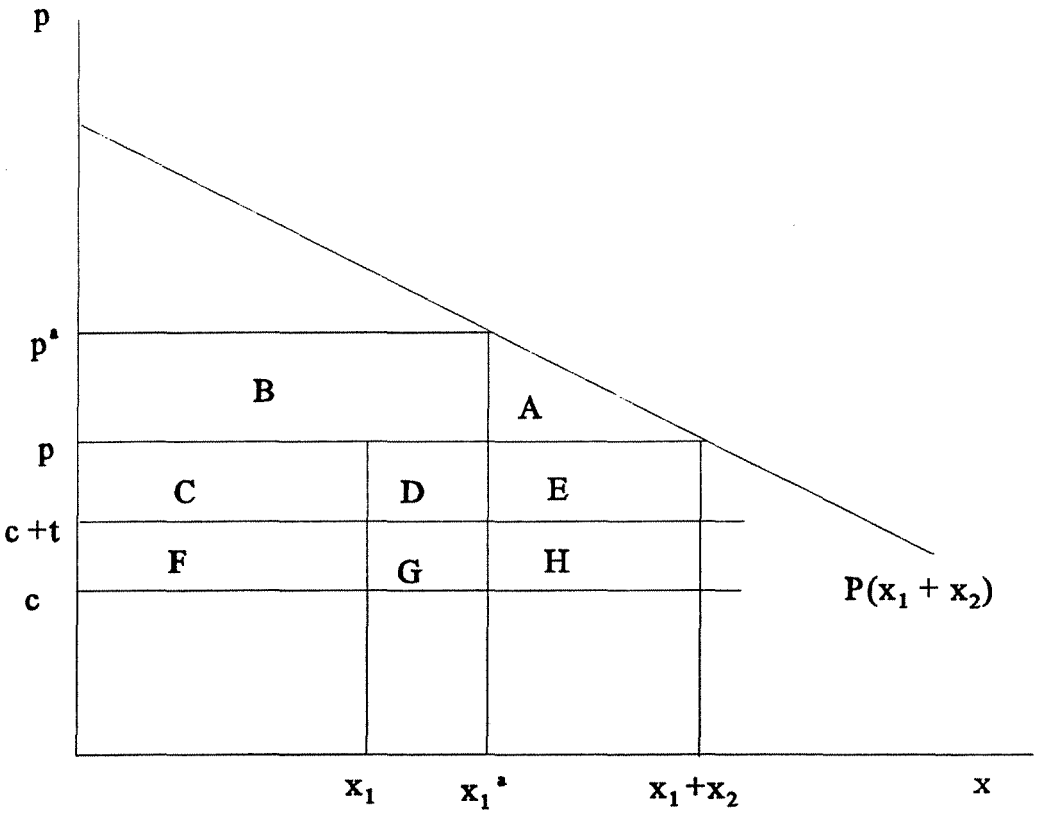

Figure 2.3

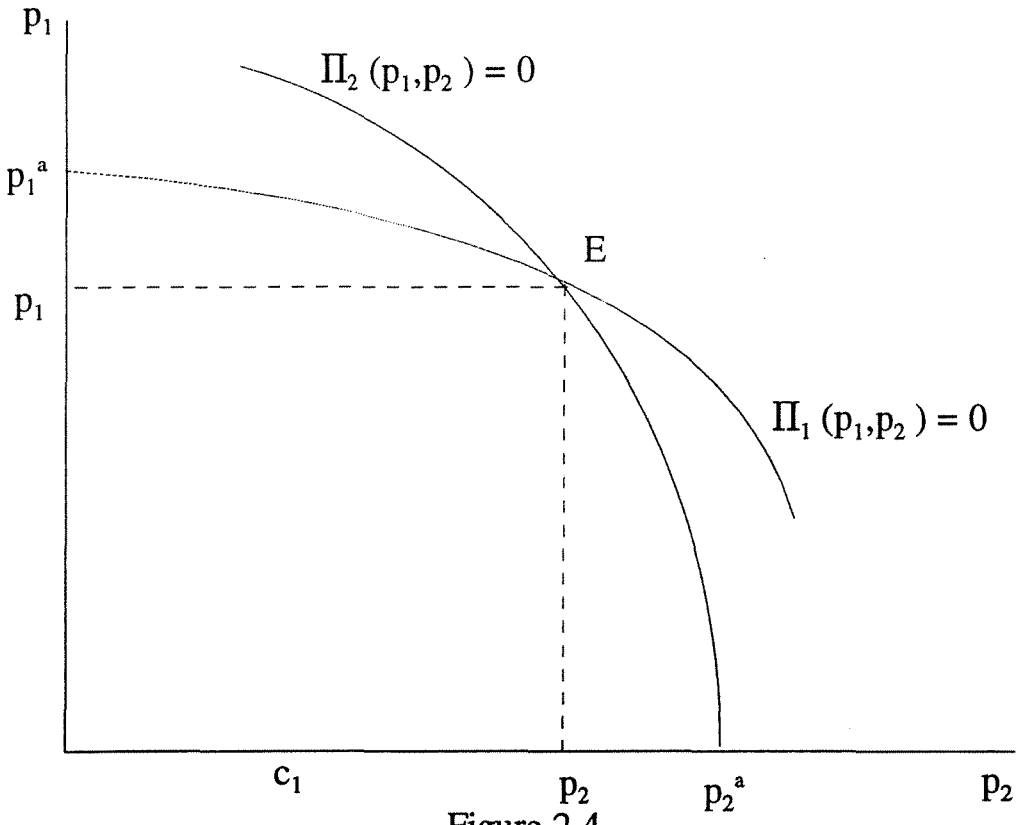

Figure 2.4

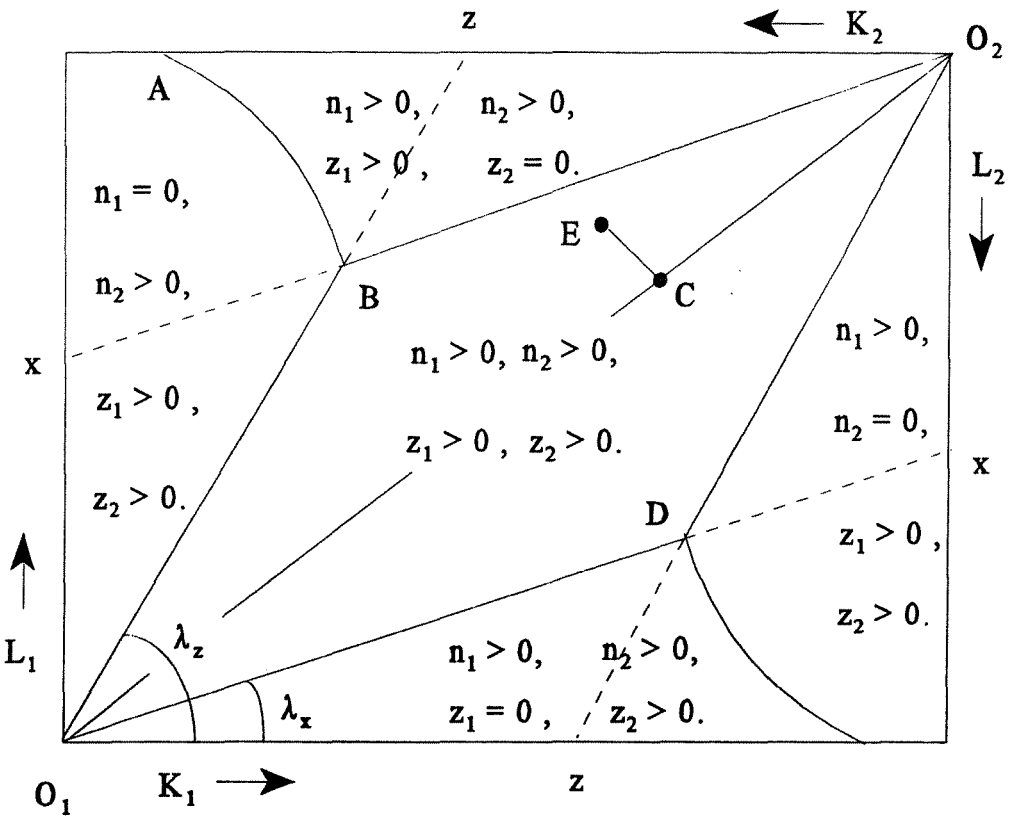

Figure 3.1: t = 1.0

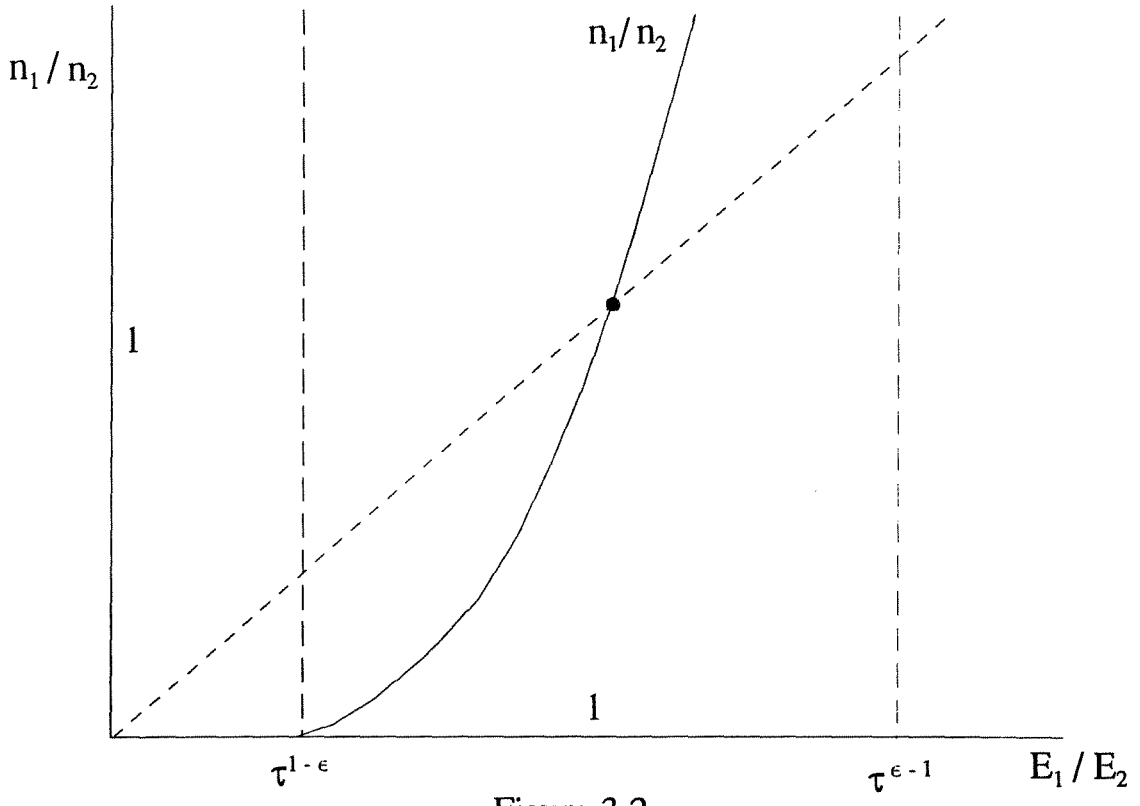

Figure 3.2

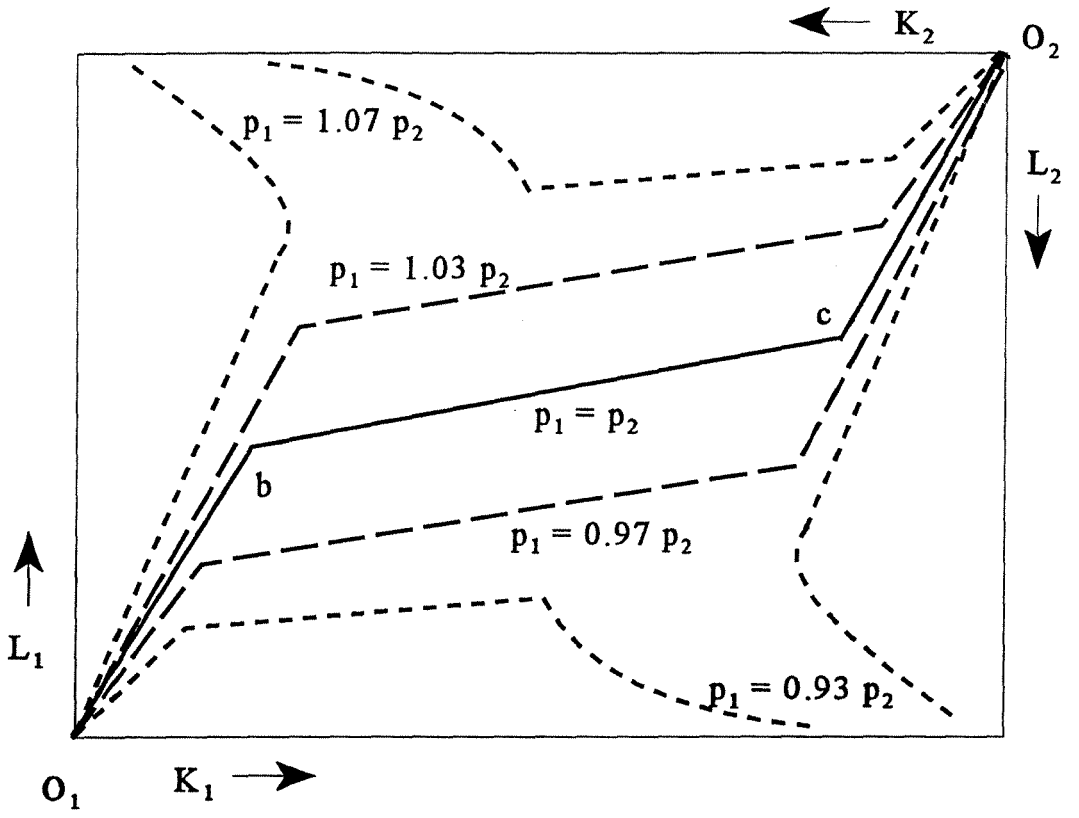

Figure 3.3: t = 1.2

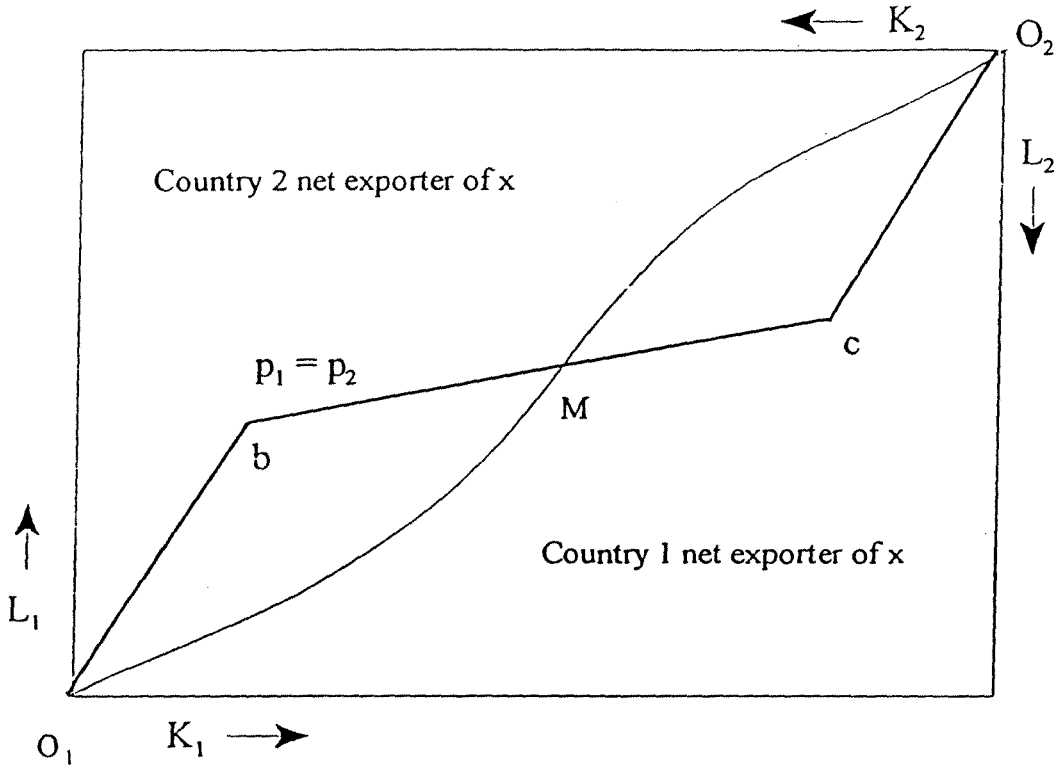

Figure 3.4: t = 1.2

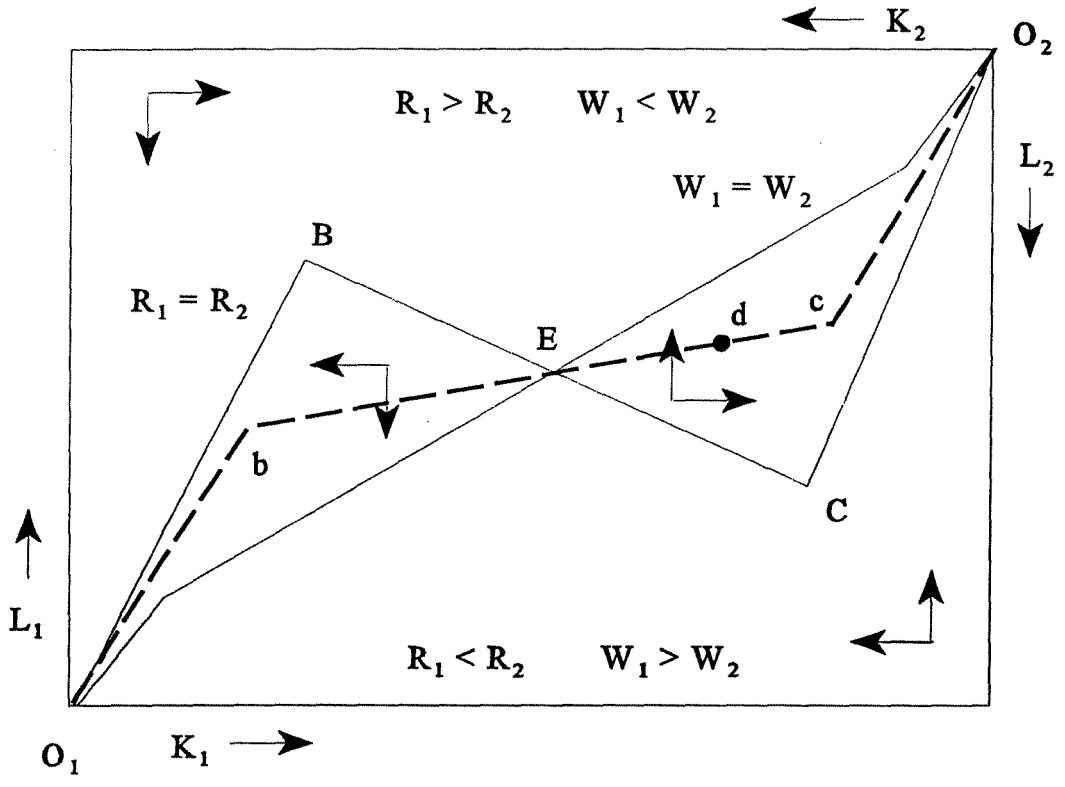

Figure 3.5

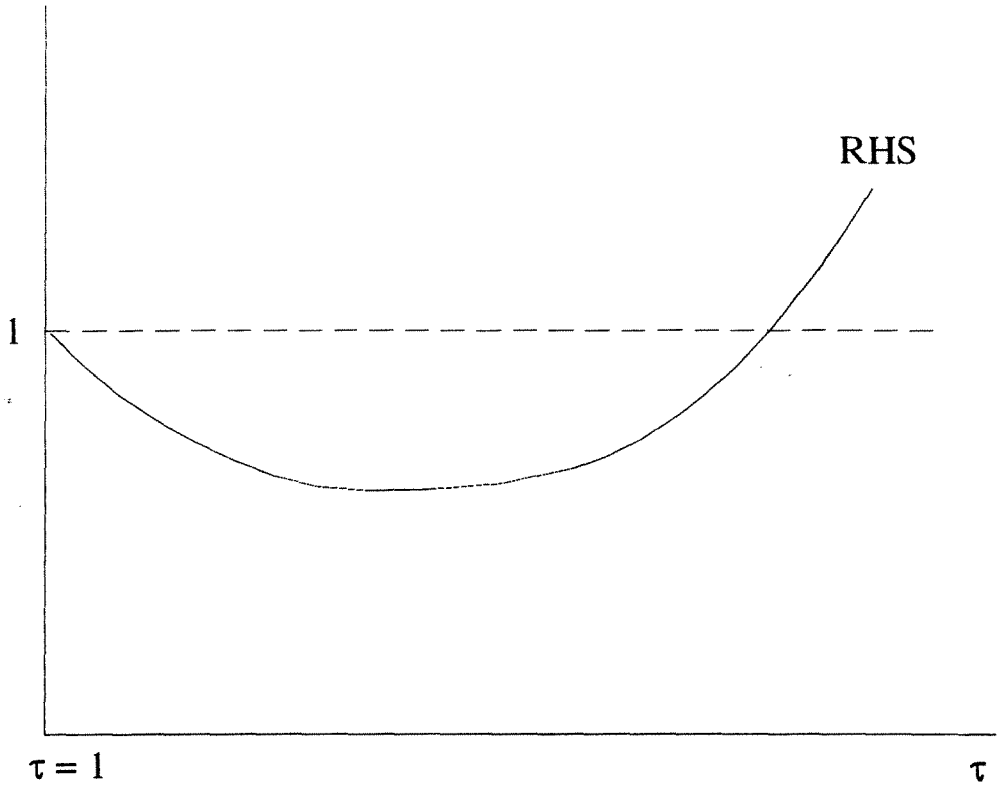

Figure 3.6