



The theory of economic integration

ALI EL-AGRAA

6.1 Introduction

In reality, some existing schemes of economic integration, especially the EU, were either proposed or formed for political reasons, even though the arguments popularly put forward in their favour were expressed in terms of possible economic gains. However, no matter what the motives for economic integration are, it is still necessary to analyse the economic implications of such geographically discriminatory groupings; that is why I included political unions as schemes of economic integration in Chapter 1 (see Section 1.2, page 1).

The chapter begins with a static analysis of the effects of economic integration on trade and production, first in partial- and then in general-equilibrium terms; static in the sense of immediate effects that do not allow for changes in consumption and production patterns. It goes on to examine these effects in dynamic terms, allowing time for changes to occur in the consumption and production patterns. Domestic distortions in the markets are then included, followed by the incorporation of changes in international prices on the analysis. Then elements of factor mobility are ushered in before various other considerations are briefly dealt with and conclusions stated. It should be stressed, however, that this chapter requires an understanding of trade theory, but the basic concepts involved have already been introduced in the introduction to this part of the book.

To understand these effects, one needs to appreciate the *possible* sources of economic gain from economic integration. At the customs union (CU) and free trade area (FTA) level they can be attributed to:

1. enhanced efficiency in production made possible by increased specialization in accordance with the law of comparative advantage;

2. increased production level due to better exploitation of economies of scale made possible by the increased size of the market;
3. an improved international bargaining position, made possible by the larger size, leading to better terms of trade (t/t);
4. enforced changes in economic efficiency brought about by enhanced competition;
5. changes affecting both the amount and quality of the factors of production arising from technological advances.

If the level of economic integration is to proceed beyond the CU level to the economic union level, then further sources of gain become *possible* as a result of:

6. factor mobility across the borders of member nations;
7. the coordination of monetary and fiscal policies;
8. the goals of near full employment, higher rates of economic growth and better income distribution becoming unified targets.

I shall now discuss these considerations in some detail.

6.2 The customs union aspects

6.2.1 The basic concepts

Before the theory of second best was introduced, it used to be the accepted tradition that CU formation should be encouraged. The rationale for this was that since free trade maximized world welfare, and since CU formation was a move towards free trade, CUs increased welfare, even though they did not maximize it. This rationale certainly lies behind the guidelines of the GATT-WTO Article XXIV (see Section 1.3, page 2), which permits the formation of CUs and FTAs as the

special exceptions to the rules against international discrimination.

Viner (1950), and arguably Byé (1950), challenged this proposition by stressing that CU formation is by no means equivalent to a move to free trade, since it amounts to free trade between the members and protection vis-à-vis the outside world. This combination of free trade and protectionism could result in trade creation and/or trade diversion. Trade creation (TC) is the replacement of expensive domestic production by cheaper imports from a partner, and trade diversion (TD) is the replacement of cheaper initial imports from the outside world by more expensive imports from a partner. Viner stressed that TC is beneficial since it does not affect the rest of the world, while TD is harmful; it is the relative strength of these two effects that determines whether or not CU formation should be advocated. It is therefore important to understand the implications of these concepts.

Assuming perfect competition in both the commodity and factor markets, automatic full employment of all resources, costless adjustment procedures, perfect factor mobility nationally, but perfect immobility across national boundaries, prices determined by cost, three countries, H (the home country), P (the potential CU partner) and W (the outside world), plus all the traditional assumptions employed in tariff theory, we can use a simple diagram to illustrate these two concepts.

In Figure 6.1, I use partial equilibrium diagrams, but will employ general equilibrium ones in most of the rest of this chapter, even though it has been demonstrated that partial and general equilibrium analyses are, under certain circumstances, equivalent (see El-Agraa and Jones 1981). S_W is W 's perfectly elastic tariff-free supply curve, for this commodity; S_H is H 's supply curve, while $S_{H,P}$ is the joint H and P tariff-free supply curve. With a non-discriminatory tariff (t) imposition by H of $AD (= t_p)$, the effective supply curve facing H is $BREFQT$ - that is, its own supply curve up to E , then that of W inclusive of the tariff $[SW(1 + t_p)]$; obviously $S_{H,P}(1 + t_p)$ will lie above S_H and hence would be out of the picture. The domestic price is therefore OD , which gives domestic production of Oq_2 , domestic consumption of Oq_3 and imports of q_2q_3 . H pays $q_2LMq_3 (= a)$ for the imports, while the domestic consumer pays $q_2EFq_3 (= a + b + c)$, with the difference ($LEFM = b + c$) being the tariff revenue which accrues to the H government. This government revenue can be viewed as a transfer from

the consumers to the government, with the implication that, when the government spends it, the marginal valuation of that expenditure should be exactly equal to its valuation by the private consumers, so that no distortions should occur.

If H and W form a CU, the free trade position will be restored, so that Oq_5 will be consumed in H and this amount will be imported from W . Hence free trade is obviously the ideal situation. But if H and P form a CU, the tariff imposition will still apply to W while it is removed from P . The effective supply curve in this case is $BRGQT$. The union price falls to OC , resulting in a fall in domestic production to Oq_1 , an increase in consumption to Oq_4 and an increase in imports to q_1q_4 . These imports now come from P .

The welfare implications of these changes can be examined by employing the concepts of consumers' and producers' surpluses. As a result of increased consumption, consumers' surplus rises by $CDFG (= d + e + c + f)$. Part of this (d) is a fall in producers' surplus due to the decline in domestic production, and another part (c) is a portion of the tariff revenue now transferred back to the consumer subject to the same condition of equal marginal valuation. This leaves e and f as gains from CU formation. However, before we conclude whether or not these triangles represent net gains, we need to consider the overall effects more carefully.

The fall in domestic production from Oq_2 to Oq_1 leads to increased imports of q_1q_2 . These cost q_1JLq_2 to import from P , while they originally cost q_1JLq_2 to produce domestically. (Recall the assumption that these resources are to be employed elsewhere in the economy without any adjustment costs or redundancies.) There is therefore a saving of e . The increase in

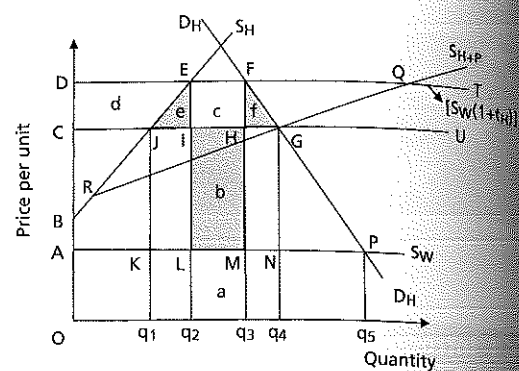


Figure 6.1 Trade creation and trade diversion

the consumers to the government, with the implication that, when the government spends it, the marginal valuation of that expenditure should be exactly equal to its valuation by the private consumers, so that no distortions should occur.

If H and W form a CU, the free trade position will be restored, so that Oq_5 will be consumed in H and this amount will be imported from W . Hence free trade is obviously the ideal situation. But if H and P form a CU, the tariff imposition will still apply to W while it is removed from P . The effective supply curve in this case is $BRGQT$. The union price falls to OC , resulting in a fall in domestic production to Oq_1 , an increase in consumption to Oq_4 and an increase in imports to q_1q_4 . These imports now come from P .

The welfare implications of these changes can be examined by employing the concepts of consumers' and producers' surpluses. As a result of increased consumption, consumers' surplus rises by $CDFG (= d + e + c + f)$. Part of this (d) is a fall in producers' surplus due to the decline in domestic production, and another part (c) is a portion of the tariff revenue now transferred back to the consumer subject to the same condition of equal marginal valuation. This leaves e and f as gains from CU formation. However, before we conclude whether or not these triangles represent net gains, we need to consider the overall effects more carefully.

The fall in domestic production from Oq_2 to Oq_1 leads to increased imports of q_1q_2 . These cost q_1JLq_2 to import from P , while they originally cost q_1JLq_2 to produce domestically. (Recall the assumption that these resources are to be employed elsewhere in the economy without any adjustment costs or redundancies.) There is therefore a saving of e . The increase in

consumption from Oq_3 to Oq_4 leads to new imports of q_3q_4 , which cost q_3HGq_4 to import from P . These give a welfare satisfaction to the consumer equal to q_3FGq_4 . There is therefore an increase in satisfaction of f . However, the initial imports of q_2q_3 cost the country a , but these imports now come from P , costing $a + b$. Therefore these imports lead to a loss in government revenue of b (c being a retransfer). It follows that the triangle gains ($e + f$) have to be compared with the loss of tariff revenue (b) before a definite conclusion can be made regarding whether or not the net effect of CU formation has been one of gain or loss.

It should be apparent that q_2q_3 represents, in terms of our definition, TD, and $q_1q_2 + q_3q_4$ represents TC, or, alternatively, that areas $e + f$ are TC (benefits), while area b is TD (loss). (The reader should note that I am using Johnson's (1974) definition so as to avoid the unnecessary literature relating to a trade-diverting welfare-improving CU promoted by Gehrels (1956-7), Lipsey (1960) and Bhagwati (1971).) It is obvious, then, that TC is economically desirable, while TD is undesirable: hence Viner's conclusion that it is the relative strength of these two effects that should determine whether or not CU formation is beneficial or harmful.

The reader should note that if the initial price is that given by the intersection of D_H and S_H (due to a higher tariff rate), the CU would result in pure TC, since the tariff rate is prohibitive. If the price is initially OC (due to a lower tariff rate), then CU formation would result in pure TD. It should also be apparent that the size of the gains and losses depends on the price elasticities of S_H , $S_{H,P}$ and D_H , and on the divergence between S_W and $S_{H,P}$ - that is, international cost differences.

6.2.2 The Cooper-Massell criticism

Viner's conclusion was challenged by Cooper and Massell (1965a). They suggested that the reduction in price from OD to OC should be considered in two stages: first, by reducing the tariff level indiscriminately - that is, for both W and P - to AC , which gives the same union price and production, consumption and import changes; second, by introducing the CU starting from the new price OC . The effect of these two steps is that the gains from the TC ($e + f$) still accrue, while the losses from TD (b) no longer apply, since the new effective supply curve facing H is $BIGU$, which ensures that imports continue to come from W at the

cost of a . In addition, the new imports due to TC ($q_1q_2 + q_3q_4$) now cost less, leading to a further gain of $KJIL$ plus $MHGN$. Cooper and Massell then conclude that a policy of unilateral tariff reduction (UTR) is superior to customs union formation. This criticism was challenged by Wonnacott and Wonnacott (1981), but their position was questioned by El-Agraa and Jones (2000a, b), although El-Agraa (2002a) demonstrates that it can be validated when WTO's Article XXIV rules are incorporated into the analysis; I shall return to these considerations in Section 6.2.7, since a different theoretical model is needed for these analyses.

6.2.3 Further contributions

Immediately following the Cooper-Massell criticism came two independent but somewhat similar contributions to the theory of CUs. The first development was by Cooper and Massell (1965b) themselves, the essence of which is that two countries acting together can do better than if each acts in isolation. The second was by Johnson (1965b), and was a private plus social costs and benefits analysis expressed in political economy terms. Both contributions utilize a 'public good' argument, with Cooper and Massell's expressed in practical terms and Johnson's in theoretical terms. However, because the Johnson approach is expressed in familiar terms, this section is devoted to it, since space limitations do not permit a consideration of both. There is, however, another reason for doing so: most of the new developments mentioned later can be tackled within this framework.

Johnson's method is based on four major assumptions:

1. Governments use tariffs to achieve certain non-economic (political, etc.) objectives.
2. Actions taken by governments are aimed at offsetting differences between private and social costs. They are, therefore, rational efforts.
3. Government policy is a rational response to the demands of the electorate.
4. Countries have a preference for industrial production.

In addition to these assumptions, Johnson makes a distinction between private and public consumption goods, real income (utility enjoyed from both private and public consumption, where consumption is the

sum of planned consumption expenditure and planned investment expenditure) and real product (defined as total production of privately appropriable goods and services).

These assumptions have important implications. First, competition among political parties will make the government adopt policies that will tend to maximize consumer satisfaction from both 'private' and 'collective' consumption goods. Satisfaction is obviously maximized when the *rate of satisfaction per unit of resources is the same in both types of consumption goods*. Second, collective preference for industrial production implies that consumers are willing to expand industrial production (and industrial employment) beyond what it would be under free international trade.

Tariffs are the main source of financing this policy, simply because GATT-WTO regulations rule out the use of export subsidies, and domestic political considerations make tariffs, rather than the more efficient production subsidies, the usual instruments of protection.

Protection will be carried to the point where the *value of the marginal utility derived from collective consumption of domestic and industrial activity is just equal to the marginal excess private cost of protected industrial production*.

The marginal excess cost of protected industrial production consists of two parts: the marginal production cost and the marginal private consumption cost. The marginal production cost is equal to the proportion by which domestic cost exceeds world market costs. In a very simple model this is equal to the tariff rate. The marginal private consumption cost is equal to the loss of consumer surplus due to the fall in consumption brought about by the tariff rate that is necessary to induce the marginal unit of domestic production. This depends on the tariff rate and the price elasticities of supply and demand.

In equilibrium, the proportional marginal excess private cost of protected production measures the marginal 'degree of preference' for industrial production. This is illustrated in Figure 6.2, where S_w is the world supply curve at world market prices; D_H is the constant-utility demand curve (at free trade private utility level); S_H is the domestic supply curve; S_{H+u} is the marginal private cost curve of protected industrial production, including the excess private consumption cost (FE is the first component of marginal excess cost

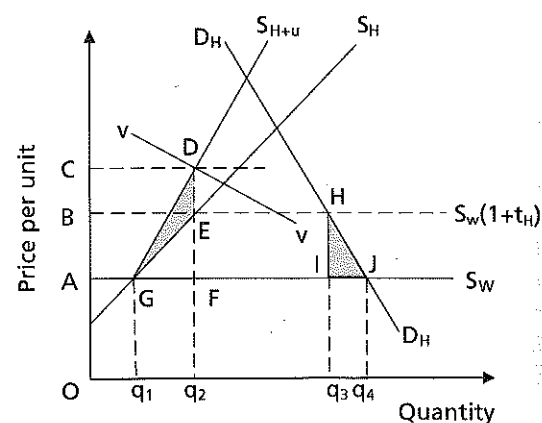


Figure 6.2 Marginal 'degree of preference' for industrial production

– determined by the excess marginal cost of domestic production in relation to the free trade situation due to the tariff imposition (AB) – and the area GED ($= IHI$) is the second component, which is the dead loss in consumer surplus due to the tariff imposition); the height of vv above S_w represents the marginal value of industrial production in collective consumption; and vv represents the preference for industrial production that is assumed to yield a diminishing marginal rate of satisfaction.

The maximization of *real* income is achieved at the intersection of vv with S_{H+u} requiring the use of tariff rate AB/OA to increase industrial production from Oq_1 to Oq_2 and involving the marginal degree of preference for industrial production v . Note that the higher the value of v , the higher the tariff rate, and that the degree of protection will tend to vary inversely with the ability to compete with foreign industrial producers. It is also important to note that, in equilibrium, the government is maximizing real income, not real product: maximization of real income makes it necessary to sacrifice real product in order to gratify the preference for collective consumption of industrial production. It is also important to note that this analysis is not confined to net importing countries. It is equally applicable to net exporters, but lack of space prevents such elaboration (see El-Agraa 1984b for a detailed explanation).

The above model helps to explain the significance of Johnson's assumptions. It does not, however, throw any light on the CU issue. To make the model useful for this purpose it is necessary to alter some of the

assumptions. Let us assume that industrial production is not one aggregate, but a variety of products in which countries have varying degrees of comparative advantage, that countries differ in their overall comparative advantage in industry as compared with non-industrial production, that no country has monopoly-monopsony power (conditions for optimum tariffs do not exist) and that no export subsidies are allowed (GATT-WTO).

The variety of industrial production allows countries to be both importers and exporters of industrial products. This, in combination with the preference for industrial production, will motivate each country to practise some degree of protection.

Given the third assumption, a country can gratify its preference for industrial production only by protecting the domestic producers of the commodities it imports (import-competing industries). Hence the condition for equilibrium remains the same: $vv = S_{H+u}$. The condition must now be reckoned differently, however: S_{H+u} is slightly different because, first, the protection of import-competing industries will reduce exports of both industrial and non-industrial products (for balance of payments purposes). Hence, in order to increase total industrial production by one unit, it will be necessary to increase protected industrial production by more than one unit so as to compensate for the induced loss of industrial exports. Second, the protection of import-competing industries reduces industrial exports by raising their production costs (because of perfect factor mobility). The stronger this effect, *ceteris paribus*, the higher the marginal excess cost of industrial production. This will be greater the larger the industrial sector compared with the non-industrial sector and the larger the protected industrial sector relative to the exporting industrial sector.

If the world consists of two countries, one must be a net exporter and the other necessarily a net importer of industrial products, and the balance of payments is settled in terms of the non-industrial sector. Therefore for each country the prospective gain from reciprocal tariff reduction must lie in the expansion of exports of industrial products. The reduction of a country's own tariff rate is therefore a source of loss, which can be compensated for only by a reduction of the other country's tariff rate (for an alternative, orthodox, explanation, see El-Agraa 1979b, c).

What if there are more than two countries? If reciprocal tariff reductions are arrived at on a most-favoured

nation basis, then the reduction of a country's tariff rate will increase imports from *all* the other countries. If the tariff rate reduction is discriminatory (starting from a position of non-discrimination), however, then there are two advantages: first, a country can offer its partner an increase in exports of industrial products without any loss of its own industrial production by diverting imports from third countries (TD); second, when TD is exhausted, any increase in partner industrial exports to this country is exactly equal to the reduction in industrial production in the same country (TC), thus eliminating the gain to third countries.

Therefore, discriminatory reciprocal tariff reduction costs each partner country less, in terms of the reduction in domestic industrial production (if any) incurred per unit increase in partner industrial production, than does non-discriminatory reciprocal tariff reduction. On the other hand, preferential tariff reduction imposes an additional cost on the tariff-reducing country: the excess of the costs of imports from the partner country over their cost in the world market.

The implications of this analysis are as follows:

1. Both TC and TD yield a gain to the CU partners.
2. TD is preferable to TC for the preference-granting country, since a sacrifice of domestic industrial production is not required.
3. Both TC and TD may lead to increased efficiency due to economies of scale.

Johnson's contribution has not been popular because of the nature of his assumptions. His economic rationale for CUs, resting on public goods grounds, can only be established if for political or similar reasons governments are denied the use of direct production subsidies. While this may be the case in certain countries at certain periods in their economic evolution, there would appear to be no acceptable reason why this should generally be true. Johnson's analysis demonstrates that CUs and other acts of commercial policy 'may make economic sense under certain restricted conditions, but in no way does it establish or seek to establish a general argument for these acts' (Krauss 1972).

6.2.4 General equilibrium analysis

The conclusions of the partial equilibrium analysis can easily be illustrated in general equilibrium terms. To simplify the analysis we shall assume that H is a 'small'