

Chapter 2

The Theory of Trade Agreements, Economic Integration, Size of Economies, Trade Costs and Welfare

2.1 Preferential Trade Agreements

Preferential trade agreements (PTAs) reduce or eliminate policy-imposed barriers to the flow of goods, services, capital, labour etc. (Baier et al. 2008). It is possible to view PTAs from three separate perspectives: economics, politics and commitment or credibility (Bagwell and Staiger 2002). Economics explains PTAs as restricting the setting of unilateral trade policy, which can affect the welfare of other nations. The political perspective views PTAs as limiting the distributional issues of trade policy choices by government. The commitment perspective explains PTAs in the context of increasing the credibility of government actions. All of these perspectives have a similar objective: to maximise national welfare, subject to different constraints. The WTO, formerly the GATT, has the principal economic objective of solving the central problem of the terms of trade externalities that arise when countries make trade policy decisions unilaterally. In so doing, the WTO applies rules that have the principles of reciprocity and non-discrimination as their cornerstone.

The main objective of this section is to provide a brief review of the theory and framework of PTAs. It examines the definition of PTAs, the GATT/WTO rules with respect to PTAs, the different stages of economic integration agreements and the composition of major economic integration agreements.

2.1.1 *PTAs and Regional Trade Agreements*

Trade liberalisation can take several forms. The simplest path to liberalisation is for a nation to cut tariffs unilaterally but, more usually, nations lower their import barriers at the same time as their trade partners. Such liberalisation can be in the form of a multilateral agreement – such as the various GATT/WTO Rounds – or an agreement among a smaller subset of nations. This latter type of agreement is

referred to as a preferential trade agreement (PTA). The defining characteristic of a PTA is that lower tariffs are imposed on goods produced in the Member countries than on those produced outside (Panagariya 2000). This subset of nations often has some geographical proximity such that is usually referred to as a regional trade agreement (RTA).

Nearly all PTAs are regional treaties between nations to reduce or eliminate policy-imposed trade barriers. Formal RTAs may cover a spectrum of arrangements, from small margins of tariff preference to full scale economic integration. All PTAs or RTAs fall within the scope of GATT 1994 Article XXIV. This Article exempts Member states from the Most-Favoured Nation (MFN) principle by treating mutual imports preferentially through the formation of a PTA. The WTO permits three types of PTA:

- Custom unions and free trade agreements sanctioned under Article XXIV.
- Agreements between developing states formed under the Enabling Clause that allows partial preferential treatment.
- Agreements under the Generalised System of Preferences (GSP) that allow developed states to grant preferential treatment to developing states.

Nations that are geographically proximate should be encouraged to form PTAs while those that are more distant should be discouraged. This argument has two elements. First, trade blocs are more likely to create trade and thus raise welfare, given a country's volume of international trade; the higher is the proportion of trade with the country's trade bloc partners and the lower is this proportion with the non-member countries. Second, countries that share borders or are closer geographically to one another tend to have a higher proportion of trade with one another than those countries that are further apart and so are more likely to be trade creating as opposed to trade diverting.

Bhagwati argues that the first argument is not valid unless substitution between commodities is considered (Bhagwati 1993). Lipsey focuses on the relative volume of imports from each source *vis-à-vis* expenditure on domestic goods as the decisive factor in determining the size of gains and losses from preferential cuts in trade barriers (Lipsey 1958). It is therefore important to look at the estimates of substitution elasticities among goods as well as trade shares with and between members and non-members. A prime criticism of the second argument is that borders can breed hostility and may undermine trade, just as alliances among distant countries with a shared cause can promote trade (Gowa and Mansfield 1994).

With respect to the formation of customs unions (CUs) or free trade areas (FTAs), El-Agraa (1994) attributes the principal potential sources for economic gain from economic integration as being:

- Enhanced efficiency in production made possible by increased specialisation.
- Increased production levels due to better exploitation of economies of scale made possible by the increased size of the market.
- An improved international bargaining position, made possible by the larger size, leading to better terms of trade.

- Enforced changes in efficiency brought about by intensified competition between firms.
- Changes affecting both the amount and quality of the factors of production due to technological advances.

Proceeding to a common market leads to further sources of gain as a result of:

- Factor mobility across the borders of member states.

Establishing an economic union also results in:

- The co-ordination of monetary and fiscal policies.

In the presence of scale economies or imperfect competition, there can be important welfare implications of forming a preferential trade area. An industry can capitalise on scale economies more easily in the larger market of a preference area. Within a larger market, firms rationalise production, produce larger output runs and effectively lower their average costs simply because a larger market without protective trade barriers is available.

The formation of preferential trade areas is generally welfare-improving when extensive imperfect competition is present in the initially protected economy. An imperfectly competitive domestic sector is forced to compete with imports such that freer trade leads to both gains from trade and increased competition and efficiency. These competitive effects relate to potential changes in production costs, reduced profit margins, the introduction of new products, increased competitive pressure on domestic producers and changes in the parameters underlying strategic decisions. The interaction of these effects with trade and trade policy can be quite complex although the minimum conditions for welfare gains are generally linked to changes in industry output.

2.1.2 The GATT and the WTO

‘If the world trading system has a constitution, it is embodied in the Articles of the General Agreement on Tariffs and Trade (GATT) and its successor organisation, the World Trade Organization (WTO)’ (Bagwell and Staiger 2002). PTAs are international trade policy agreements and are therefore governed by the GATT rules.

The WTO was established on 1 January 1995. The WTO Agreements include the text of the GATT 1994, as well a set of additional agreements. Its goal is to reduce barriers to trade among its Members: ‘it is an organisation for liberalising trade’ (WTO web-site). This has been clear in successive GATT Rounds, which have aimed at multilateral reductions in trade barriers. There are currently 153 WTO Members, the latest being Cape Verde joining in July 2008. The majority of Members are participants in RTAs; according to the WTO, there are close to 400 PTAs that are scheduled to be implemented by 2010. Of these, FTAs and partial scope agreements account for over 90 % while CUs account for less than 10 %. Prior to 1995, there were 124 RTAs in force.

Among the best known RTAs are the EU, EFTA, CISFTA, NAFTA, MERCOSUR, ASEAN and COMESA; synopsis of each are given in Chapter 8, Sect. 8.1.

The signing of the WTO Agreements in 1995 reasserted a strong integrationist focus in trade matters as Members of the new organisation henceforward were compelled to accept all of the wide ranging international arrangements. Members are no longer able to 'cherry-pick' those arrangements that suit their own interest while ignoring those that are less beneficial (Phillips 2007).

Prior to the theory of 'second best', it was an accepted economic assumption that any PTA was welfare improving. The rationale was that, since free trade maximises world welfare and preferential trade represents a move towards free trade, PTAs therefore increase welfare, although they do not maximise it. This rationale lies behind the guidelines of the original GATT 1947 Article XXIV, which permits the formation of PTAs – CUs and FTAs – as an exception to the rule against international discrimination.

The centrepiece of the GATT rules is the Most Favoured Nation (MFN) principle, Article I. According to MFN, each WTO Member grants all WTO Members the same advantage, privilege, favour, or immunity that it grants to any other state. A key implication of this provision is that WTO Member states cannot discriminate in their tariff policy between Members. The MFN principle is designed to prevent the development of bilateral preferential trade treatment under which the pattern of trade could become distorted and less than optimal. Under the MFN principle, when a Member extends trade concessions to one partner, it must extend them to all such that it is therefore a principle of non-discrimination. The only exception is the variation of MFN through the creation of a PTA under certain circumstances; for instance, if 'almost all' trade among the parties to the agreement is covered.

There exist three separate provisions for trade preferences within the GATT/WTO framework. First, developed countries can grant developing states non-reciprocal trade preferences. Second, developing countries can exchange any trade preferences to which they agree. Finally, under Article XXIV, any two or more members of the WTO can form a PTA. GATT Article XXIV offers the only avenue to PTAs in which developed states are recipients of trade preferences.

Article XXIV states that a group of two or more customs territories may form a trade bloc by reducing barriers among themselves, subject to several requirements. The first requirement is that 'substantially all' barriers among the Members are removed. According to Frankel (1997), 'substantially all' is interpreted both as liberalisation covering a high percentage of total trade – 80 % in the case of the formation of the EU in 1957 – and occurring in most major sectors, as in the formation of EFTA in 1960 (GATT 1994, pp. 738, 766–768). The second requirement is that trade barriers against non-members are not made more restrictive than before. When Members enter a PTA with different tariffs against non-members, the new external tariffs must be no more than the weighted average of the pre-integration tariffs. If the net effect is to raise barriers in some sectors, affected non-members may claim compensation. Finally, progress toward economic integration is supposed to be expeditious, normally not to exceed 10 years (as defined in the Uruguay Round negotiations in 1994).

2.2 The Theory of Economic Integration

Traditional economic theory asks two questions relating to trade liberalisation and the formation of a preferential trade agreement (PTA) – whether a free trade area (FTA) or a customs union (CU) – what is the welfare impact on each member country, the bloc as a whole and the rest of the world?; and can two or more countries form a trade bloc such that it makes the rest of the world worse off? These two questions remain as valid today as 60 years ago.

The first question is investigated by Viner (1950), who introduced the concepts of trade creation and trade diversion. Meade (1955) provides the first welfare theoretic analysis of trade blocs in a general equilibrium model, which focuses on the latter question. Meade's model has since been extended to answer both questions by significant contributions from Lipsey (1958), Mundell (1964), Vanek (1965), Corden (1976) and McMillan and McCann (1981). Of these contributions, only Viner and Vanek explicitly distinguish trade blocs as CUs involving internal free trade and a common external tariff. Almost all other contributors use the term custom union more loosely, involving internal free trade but members retaining their original tariff levels against non-members. With this in mind, the term trade bloc is used synonymously for custom unions and free trade areas. A trade bloc is defined as an agreement between nations to eliminate trade policy barriers on goods among participating states.

The main objective of this section is to discuss the theory of economic integration and to provide a theoretical background to the empirical analysis undertaken. It examines the definition of economic integration, discusses traditional welfare analysis and general equilibrium analysis and establishes a trade flow measure for theoretical measurement of welfare and comparison.

2.2.1 *Defining Economic Integration*

There is no clear-cut definition of economic integration in the international trade literature. Its objective is clear; it is a means to increase welfare. Its vague definition however, implies that there is no general agreement regarding the method to achieve this goal. Balassa (1962) defines economic integration as both a process and a state of affairs. Integration is a process in that involves the removal of trade discrimination between different states, while it is a state of affairs to the extent that it is the absence of different forms of discrimination. Robson (1987) refers to economic integration as being basically concerned with efficiency in resource use, with particular reference to spatial aspects. The necessary conditions for its fullest attainment include the freedom of movement of goods and factors of production and an absence of discrimination amongst members. In addition, where resources are allocated by the price mechanism, measures are required to ensure that the market provides the right signals and institutions are required to give effect to the integrating force of the market.

Molle (1990) takes economic integration to indicate the gradual elimination of economic frontiers between countries. In the first stage, trade among partners is liberalised. This is followed by the liberalisation of movement of production factors. The objective of the third stage is the co-ordination of national policies with regard to economic sectors, possibly including exchange rates. El-Agraa (1994) refers to economic integration as the discriminatory removal of all trade impediments between participating nations and the establishment of certain elements of co-operation and co-ordination between them. Pelkmans (2006) follows Molle in viewing economic integration as the elimination of economic frontiers between two or more economies. An economic frontier is any demarcation over which actual and potential mobility of goods, services and production factors, as well as communication flows, are relatively low. This book defines international economic integration as a process of eliminating trade cost such that it is a means to reduce trade costs to increase welfare. As a process, it is evolving and continuing with changes in markets.

2.2.2 The Traditional Welfare Analysis of Economic Integration

Until Viner's penetrating analysis in 1950, both free traders and protectionists argued in favour of trade blocs. The former saw only the benefits of free intra-bloc trade while the latter emphasised the benefits of protection from non-members' goods. Viner's introduction of the key concepts of trade creation and trade diversion however, demonstrated that trade blocs were not necessarily welfare improving, whether for Member states or globally, such that trade blocs might harm welfare.

Viner associates trade creation with a welfare gain and trade diversion with a welfare loss. Whether or not a trade bloc is welfare increasing depends upon the relative magnitudes of trade creation and trade diversion. Trade creation is the replacement of domestic production by lower cost imports from a partner and trade diversion is the replacement of lower cost cheaper imports from the world market by more expensive imports from a partner. Viner stresses that trade creation is beneficial, since it does not affect the rest of the world, while trade diversion is harmful. It is therefore the relative strength of these two effects that determines whether or not a trade bloc is welfare enhancing or not (Viner 1950).

Bhagwati and Panagariya (1996) point out that conventional trade creation and trade diversion are not the entire story in deciding on the welfare outcome for an individual member of a trade bloc. Even if trade creation is larger than trade diversion, so that the bloc as a whole benefits, an individual member could lose on account of adverse income distribution effects arising from tariff revenue redistribution. This implies that, when an economy with a high degree of protection forms a trade bloc with an economy with relatively open markets, the former may well be faced with a net welfare loss. Trade diversion can also, under certain circumstances, be beneficial; for example, if a member country introduces imports

into the domestic market that reduce distortions in consumers' patterns of consumption (Meade 1955; Gehrels 1956–1957; Lipsey 1957). Further, if economies of scale are present that allow production at a lower cost (Corden 1972; Venables 1987) or when new competition reduces the market power of inefficient domestic monopolies.

2.3 The General Equilibrium Analysis of Economic Integration

Trade costs became a major factor in the welfare analysis of the effect of trade blocs on world welfare only after Meade's influential general equilibrium analysis (Meade 1955). Meade points out that the relative magnitudes of trade creation and trade diversion alone are insufficient to determine the welfare effect of a bloc on world welfare because the benefits of preferential liberalisation depend not only upon the extent of trade creation but also on trade costs. Similarly, losses are determined not just by the amount of trade diversion but also the magnitude of the increase in costs due to trade diversion (Meade 1955).

2.3.1 *The Theory of Second Best*

The general theorem of the second best states that if a constraint is introduced into a general equilibrium setting which prevents the attainment of one of the Pareto conditions, other Pareto conditions, although still attainable are in general, no longer desirable (Lipsey and Lancaster 1956–1957). Adam Smith and David Ricardo view free trade and the unimpeded movement of factors as the first best policy in a world which does not have any distortions. Attainment of the Pareto optimum requires the simultaneous fulfilment of all optimum conditions. An allocation of resources is said to be Pareto optimal if there does not exist another feasible allocation in which some agents would be better off (in a welfare sense) and no agents worse off. Pareto optimality is achieved exclusively under free trade such that other cases where there are distortions – e.g. tariffs, subsidies, taxes, monopolies etc. – are sub-optimal.

Before the theory of second best, trade blocs were considered to be a move closer to free trade and therefore welfare increasing. The theorem of second best addresses this by stating that, in the presence of distortions, if all the conditions for Pareto optimality cannot be satisfied, then the removal of some of the distortions does not necessarily increase welfare, nor does the addition of other distortions necessarily decrease it. One sub-optimal situation is therefore replaced by another sub-optimal situation. Welfare may remain unaffected, increased or decreased. In a system with several distortions, the removal of any single distortion cannot be presumed to be welfare improving. In other words, if an economy is prevented from attaining all the conditions for maximum welfare simultaneously, the fulfilment of one of these

conditions will not necessarily make the economy better off. This is the general theorem of second best.

From this theorem, welfare comparisons between economic states are ambiguous when some Pareto optimum conditions are met while others are not. Welfare comparisons need to be done so that no one is left worse off.

2.3.2 *The Transfer Payment Principle*

Meade's model is concerned with world welfare; if lump sum transfers could be deployed, it would be possible to trace out the entire Pareto efficient frontier for all the consumers in the world. This is not realistic; it needs lump sum transfers at an international level. The first step towards realism is to use lump sum transfer within a single economy as a method to evaluate the desirability of two economic situations.

Following Ohyama (1972), a government has the ability to tax the gainers and transfer income to the losers, i.e. without changing their behaviour in the process. Lump sum transfers are assumed to be non-distorting. The analysis makes use of the traditional terms of trade definition of a small economy with a perfectly competitive market structure and constant returns to scale.

There are n commodities, some of which are final goods and some of which are intermediate goods; $\mathbf{N}$ denotes the vector¹ of commodities and $\mathbf{P}$ denotes the vector of domestic prices. The economy consists of agents, the role of each of them being to choose a complete plan of action. Each agent is characterised by the limitations on their choice and by their choice criteria. There are three distinct classes of economic agents in the economy, producers, consumers and the government. A producer is supposed to carry out a production plan which is a specification of the quantities of their inputs and outputs. The production plan is constrained to belong to a given set representing essentially their limited technological knowledge. In that set the production plan is chosen, for given prices, so as to maximise profit, the sum of all receipts minus the sum of all outlays. It is assumed that there is a given positive integral number k of producers and each one of them is indicated by an index $j = 1, \dots, k$. Let Y_j be the production set of the j th producer, which is closed in a n -dimensional commodity space. The set $\mathbf{Y} = \sum_{j=1}^k Y_j$ is the total production set, which describes the production possibilities of the whole economy.

On the production side, if $\mathbf{V}$ is the vector of endowments, the production possibility set can be denoted by $\Gamma(\mathbf{V})$. It is assumed that there is a given integral number h of consumers and each of them is indicated by an index $i = 1, \dots, h$. A given factor supply by the i th consumer is represented by V_i and $\mathbf{V} = \sum_{i=1}^h V_i$ is the

¹ All vectors are treated as column vectors and transposes are denoted by $\bullet$.

total factor supply. A consumer is supposed to carry out a consumption plan which is a specification of their consumption of commodities. A consumption plan is made subject to the constraint of the consumers' income composed of the value of their endowment of commodities, their share in producers' profits and their net transfer receipt. Let X_i be the consumption set of the i th consumer and his preference preordering $\prec_i$. The set $\mathbf{X} = \sum_{i=1}^h X_i$ is the total consumption set. Let U_i be the utility level of individual i and let $\mathbf{U} = \sum_{i=1}^h U_i$ denote the vector of utility levels for all individuals. Let $\mathbf{U}(\mathbf{X})$ be a well-behaved utility function, which is continuous, quasi-concave, and increasing in $\mathbf{X}$. Preferences can be represented by an increasing and strictly quasi-concave utility function so that demand is single-valued. It is also assumed that demand functions are continuous. Consumer i , has the following increasing and quasi-concave utility function:

$$U_i(X_i, V_i); \quad i = 1, \dots, h \quad (2.1)$$

There are two economic situations or states of the economy, denoted as $\{0, 1\}$, where consumer i consumes a vector X_i^0 of goods and supplies a vector V_i^0 of factors. Consumers maximize utility subject to a budget constraint:

$$\begin{aligned} &\text{Max } U_i(X_i, V_i) \\ &\text{s.t. } \mathbf{P}^0 \cdot X_i^0 \leq \mathbf{W}^0 \cdot V_i^0 \end{aligned} \quad (2.2)$$

Where $\mathbf{P}^0$ is the vector of commodity prices in situation 0 and $\mathbf{W}^0$ is the vector of factor prices in situation 0. The resulting utility for each consumer is $U_i(\mathbf{X}^0, \mathbf{V}^0)$. Total output is $\sum_{i=1}^h X_i^0 = \mathbf{Y}^0$ and total factor inputs are $\sum_{i=1}^h V_i^0 = \mathbf{V}^0$. Constant returns to scale imply:

$$\mathbf{P}^0 \cdot \mathbf{Y}^0 - \mathbf{W}^0 \cdot \mathbf{V}^0 = 0 \quad (2.3)$$

where $\mathbf{P}^0 \cdot \mathbf{Y}^0$ is the revenue to the producers and $\mathbf{W}^0 \cdot \mathbf{V}^0$ is the payment to factors.

The role of the government is threefold. First, it is assumed to tax and/or subsidise various economic activities. Secondly, it distributes income among consumers in a lump-sum fashion by changing the structure of individual shares in all income sources. For this purpose the government is able to impose personal tax-subsidy schemes on income derived from the ownership of commodity endowment, the share in profits and the net private transfer receipt. The government's net revenue (or cost) from all taxes and subsidies is assumed to be disposed of by lump-sum transfers to consumer to help achieve the purpose of income redistribution. Thirdly, the government carries out the production and consumption of commodities on its own.

The government is assumed to maximise national (social) welfare. Let T_i denote the transfer to each individual i . This may be positive or negative. The total disbursement of such transfers is $\sum_{i=1}^h T_i$ – and therefore the net revenue to the government organising such a scheme is $-\sum_{i=1}^h T_i$. The net revenue of the government needs to be non-negative in order for this system of lump sum transfers to be feasible.

Under free trade, the equilibrium commodity and factor prices are $(\mathbf{P}, \mathbf{W})$. With each individual i receiving the transfer T_i , the budget constraint is:

$$\mathbf{P} \cdot \mathbf{X}_i \leq \mathbf{W} \cdot \mathbf{V}_i + T_i \quad (2.4)$$

All individuals should be as well off as they were in situation 0, so if the price of a commodity rises from situation 0 to situation 1, the government will subsidise each individual by the price rise times the individual's consumption in situation 0. Conversely, if the earnings of a factor rise from situation 0 to situation 1, the government will tax each individual by the wage increase times the individual's factor supply in situation 0.

$$\begin{aligned} -\sum_{i=1}^h T_i &= (\mathbf{P}^0 - \mathbf{P}^1) \cdot \mathbf{X}^0 - (\mathbf{W}^0 - \mathbf{W}^1) \cdot \mathbf{V}^0 \\ &= (\mathbf{P}^0 - \mathbf{P}^1) \cdot \mathbf{Y}^0 - (\mathbf{W}^0 - \mathbf{W}^1) \cdot \mathbf{V}^0 \\ &= (\mathbf{P}^0 \cdot \mathbf{Y}^0 - \mathbf{W}^0 \cdot \mathbf{V}^0) - (\mathbf{P}^1 \cdot \mathbf{Y}^0 - \mathbf{W}^1 \cdot \mathbf{V}^0) \\ &= -(\mathbf{P}^1 \cdot \mathbf{Y}^0 - \mathbf{W}^1 \cdot \mathbf{V}^0) \\ &\geq -(\mathbf{P}^1 \cdot \mathbf{Y}^1 - \mathbf{W}^1 \cdot \mathbf{V}^1) = 0 \end{aligned} \quad (2.5)$$

Therefore, $-\sum_{i=1}^h T_i \geq 0$, so this system does not cost the government anything.

The lump sum taxes collected from those gaining from trade are more than enough to cover the subsidies to those harmed by trade. Lump sum transfers, whereby the government has the ability to tax the gainers and transfer income to the losers without changing their behaviour in the process, are assumed to be non-distorting. Under these conditions it is possible to achieve Pareto gains from trade where everybody gains.

The problem that arises when trying to implement this lump sum transfer procedure is that it needs too much information to implement. The government would need information about each individual's consumption and factor supplies. Even if this were possible, this could require very costly information gathering. To tackle this problem Dixit and Norman (1980) propose a system of commodity taxes and subsidies designed to achieve Pareto gains from trade requiring much less information. They restrict domestic policies to taxation of commodities and factors. The procedure is to hold commodity and factor prices for consumers fixed, while

commodity and factor prices for producers move freely. Consumers face the price vector $(\mathbf{P}^0, \mathbf{W}^0)$ and firms face the price vector $(\mathbf{P}^1, \mathbf{W}^1)$. Setting the vector of consumer taxes on commodities at $(\mathbf{P}^1 - \mathbf{P}^0)$ and the vector of consumer subsidies on factors at $(\mathbf{W}^1 - \mathbf{W}^0)$, the revenue collected from this system is,

$$\begin{aligned} & (\mathbf{P}^0 - \mathbf{P}^1) \cdot \mathbf{X}^0 - (\mathbf{W}^0 - \mathbf{W}^1) \cdot \mathbf{V}^0 \\ &= (\mathbf{P}^0 - \mathbf{P}^1) \cdot \mathbf{Y}^0 - (\mathbf{W}^0 - \mathbf{W}^1) \cdot \mathbf{V}^0 \\ &\geq -(\mathbf{P}^1 \cdot \mathbf{Y}^1 - \mathbf{W}^1 \cdot \mathbf{V}^1) = 0 \end{aligned} \quad (2.6)$$

Which is identical to (2.5). This system of commodity taxes and subsidies only requires information on the commodity and factor prices $(\mathbf{P}^0, \mathbf{W}^0)$ and $(\mathbf{P}^1, \mathbf{W}^1)$ compared to much more detailed information on lump sum transfer procedure.

The use of lump sum subsidies to achieve Pareto gain from trade can be readily used to compare any two trading situations. If trade liberalisation combined with lump sum transfers could make everyone better off, then this is a worthwhile trade policy change even if the transfers are not made. This statement is a derivation of Chipman's compensation principle: 'if the prospective gainers can compensate any prospective losers and leave no one worse off, the other state is to be selected; a decision is made between two economic states, one is the original state' (Chipman 1987).

2.3.3 Welfare Comparisons

The definition of economic integration describes it as a process from one economic situation to another. To be able to measure the effects of changes in trade costs through trade policy, a welfare comparison is needed. The establishment of welfare comparison in this section draws on the work of Ohyama (1972), Grinols and Wong (1991), Wong (1991) and Ju and Krishna (2000a, b, c).

A state of the economy is a specification of the action of each agent, and a state is said to be attainable if the action of each agent is possible for them and if their actions are compatible with the total resources. Given an economic situation, S , the objective is to compare from the viewpoint of consumers the welfare of the alternative situations S^0 and S^1 , where the superscripts are time periods $\{0, 1\}$. To be able to compare any two trade situations, a procedure is needed that allows the government to be able to tax the gainers and transfer income to the losers, without changing their behaviour in the process. The objective of trade is increased welfare and this, in turn, leads to welfare economics where the objective is the evaluation of the social desirability of alternative economic states. An economic state is a particular arrangement of economic activities and of the resources of the economy. The criterion used for evaluating policies by comparing welfare from a national point of view.

First, there is the social welfare approach, in which a Bergson-Samuelson social welfare function is defined in terms of the utility levels of different individuals. Secondly, there is the social utility approach, where all aggregate consumption

bundles are ranked with respect to a well-behaved social utility function. Samuelson (1956) shows that a social utility function exists if a social welfare function exists. The advantage of this approach is that the economy can be regarded as a single consumer, so axioms of revealed preference can be applied. Thirdly, there is the compensation approach, where the transition from one situation to another situation is said to be preferable if all losers can be compensated while at least one individual can be made better off.

Let $E(\mathbf{P}, \mathbf{U})$ and $R(\mathbf{P}, \mathbf{V})$ denote the standard expenditure and revenue functions. $E'_i(\mathbf{P}, U_i) = X_i(\mathbf{P}, U_i)$ is the vector of compensated demand functions (partial derivatives are denoted by ') of individual i and $R'(\mathbf{P}, \mathbf{V}) = Y(\mathbf{P}, \mathbf{V})$ is the aggregate supply vector. Define $E(\mathbf{P}, \mathbf{U})$ to be the sum of expenditure functions of all individuals:

$$E(\mathbf{P}, \mathbf{U}) = \sum_{i=1}^h E_i(\mathbf{P}, U_i) \quad (2.7)$$

is the aggregate expenditure level, while:

$$\mathbf{X}(\mathbf{P}, \mathbf{U}) = E'(\mathbf{P}, \mathbf{U}) = \sum_{i=1}^h E'_i(\mathbf{P}, U_i) = \sum_{i=1}^h X_i(\mathbf{P}, U_i) \quad (2.8)$$

is the aggregate consumption vector. Suppose that the economy moves from one situation to another. Denoting the initial situation with superscript 0 and the final situation by superscript 1. Economic situation S^0 (time period 0) and S^1 (time period 1):

$$\Delta W = E(\mathbf{P}^1, \mathbf{U}^1) - E(\mathbf{P}^1, \mathbf{U}^0) \quad (2.9)$$

Hence, if:

$$\Delta W = E(\mathbf{P}^1, \mathbf{U}^1) - E(\mathbf{P}^1, \mathbf{U}^0) \geq 0 \quad (2.10)$$

then $\mathbf{U}^1 \geq \mathbf{U}^0$. The consumption bundle chosen in period 0 is affordable at prices in period 1.

ΔW measures the difference between minimum expenditure, evaluated at $\mathbf{P}^1$ to reach a utility level of $\mathbf{U}^1$ and that to reach a utility level of $\mathbf{U}^0$. Because the expenditure function is increasing in utility, a necessary and sufficient condition for $\mathbf{U}^1 \geq \mathbf{U}^0$ is that $\Delta W \geq 0$.² This is a sufficient condition for welfare to rise as a result

² ΔW is related to the compensating and equivalent variations. Compensating variation (CV) is defined as $\Delta W_1 = E(\mathbf{P}^1, \mathbf{U}^1) - E(\mathbf{P}^1, \mathbf{U}^0)$ and equivalent variation (EV) as $\Delta W_2 = E(\mathbf{P}^0, \mathbf{U}^1) - E(\mathbf{P}^0, \mathbf{U}^0)$. Alternatively, CV is defined as $\Delta W_3 = E(\mathbf{P}^1, \mathbf{U}^0) - E(\mathbf{P}^0, \mathbf{U}^0)$ and EV as $\Delta W_4 = E(\mathbf{P}^1, \mathbf{U}^1) - E(\mathbf{P}^0, \mathbf{U}^1)$. These four measures are related, since it can be shown that $\Delta W_1 + \Delta W_3 = \Delta W_4 + \Delta W_2 = E(\mathbf{P}^1, \mathbf{U}^1) - E(\mathbf{P}^0, \mathbf{U}^0)$; (Varian 1992). In (2.9) ΔW_1 is used as a measure of welfare.

of the change in policy that moved prices from those prevailing in period 0 to those prevailing in period 1.

The gains from trade theorem states that the value of the free trade production bundle at free trade prices is greater than or equal to the value of the autarky production bundle:

$$\mathbf{P}^1 \cdot \mathbf{Y}^1 \geq \mathbf{P}^1 \cdot \mathbf{Y}^0 \quad (2.11)$$

Where the economic situation S^0 is autarky and S^1 is free trade. In autarky:

$$\mathbf{M}^0 = \mathbf{Y}^0 - \mathbf{X}^0 \quad (2.12)$$

With free trade, the balance of payment condition is:

$$\mathbf{P} \cdot \mathbf{Y} = \mathbf{P} \cdot \mathbf{X} \quad (2.13)$$

Substituting, (2.13) into (2.11):

$$\mathbf{P}^1 \cdot \mathbf{X}^1 \geq \mathbf{P}^1 \cdot \mathbf{X}^0 \quad (2.14)$$

Since $\mathbf{X}^1$ is chosen when $\mathbf{X}^0$ was available, the free trade consumption bundle is ‘revealed preferred’³ to the autarky consumption bundle.

$$E(\mathbf{P}^1, \mathbf{U}^1) \geq E(\mathbf{P}^1, \mathbf{U}^0) \quad (2.15)$$

Implying that $\mathbf{U}^1 \geq \mathbf{U}^0$. The second sufficient condition can be found by:

$$\mathbf{P}^1 \cdot [\mathbf{X}(\mathbf{P}^1, \mathbf{U}^1) - \mathbf{X}(\mathbf{P}^0, \mathbf{U}^0)] - \mathbf{P}^1 \cdot [\mathbf{Y}(\mathbf{P}^1, \mathbf{V}) - \mathbf{Y}(\mathbf{P}^0, \mathbf{V})] \quad (2.16)$$

The following relationship denotes imports:

$$\mathbf{Z}(\mathbf{P}, \mathbf{U}, \mathbf{V}) = \mathbf{X}(\mathbf{P}, \mathbf{U}) - \mathbf{Y}(\mathbf{P}, \mathbf{V}) \quad (2.17)$$

So, (2.16) can be rewritten as:

$$\mathbf{P}^1 \cdot [\mathbf{Z}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}) - \mathbf{Z}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V})] \geq 0 \quad (2.18)$$

³ The revealed-preferred approach neither presupposes a utility function nor preference ordering – it goes directly to the demand for commodities. If a certain bundle of commodities is actually purchased by a certain consumer at a certain price vector, it is supposed to, ‘reveal’ that the consumer prefers this bundle of commodities to the bundles of goods which cost less than or the same amount as the bundle purchased (Samuelson 1947).

For a small economy $\mathbf{P}^{W1} = \mathbf{P}^{W0} = \mathbf{P}^W$, hence:

$$\begin{aligned} & \mathbf{P}^1 \cdot [\mathbf{Z}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}) - \mathbf{Z}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V})] \\ &= (\mathbf{P}^W + \mathbf{T}^1) \cdot [\mathbf{Z}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}) - \mathbf{Z}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V})] \\ &= \mathbf{T}^1 \cdot [\mathbf{Z}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}) - \mathbf{Z}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V})] \end{aligned} \quad (2.19)$$

since by budget balance, the value of net trade at world price is zero. Therefore:

$$\mathbf{T}^1 \cdot [\mathbf{Z}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}) - \mathbf{Z}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V})] \geq 0 \quad (2.20)$$

is sufficient for $\mathbf{U}^1 \geq \mathbf{U}^0$ for the small economy case. Ohyama (1972) shows that (2.10) or (2.18) are sufficient for a change in tariff to be welfare-improving and, for a small economy, (2.20) is sufficient as well.

Extending the analysis to many consumers, the assumption about income distribution becomes important since the model is now concerned with many utility levels. These can be quite different if assumptions about the income distribution are not made. When the model looks at a representative consumer, it assumes that this particular consumer represents consumers in the whole economy. Ohyama (1972) assumes that income distribution is the same before and after the lump-sum transfer, so the government has the role of keeping the income distribution the same, with lump-sum procedures. That is, lump-sum transfers are assumed to be non-distorting.

Grinols and Wong (1991) assume that all individuals are given equal weight in welfare and then the condition in (2.10) is sufficient for a Pareto improvement to be possible through the use of suitable lump-sum taxes. If the economy moves from one economic situation to another, consumers can be compensated by a lump-sum procedure, given the income distribution. For a small economy, the single representative consumer method is plausible.

In a small economy model, it is useful to use the definition of the expenditure function:

$$E(\mathbf{P}, \mathbf{X}) = \mathbf{P} \cdot \mathbf{X} = \mathbf{P} \cdot \mathbf{Y} \quad (2.21)$$

And the welfare measure is:

$$\Delta W = E(\mathbf{P}^1, \mathbf{U}^1) - E(\mathbf{P}^1, \mathbf{U}^0) \quad (2.22)$$

2.3.4 Welfare Comparisons in Economic Integration

The analysis uses the welfare comparison from the previous section to measure the gain or loss from economic integration. Assume two countries, denoted by the subscripts *A* and *B*, decide to co-ordinate their policies in relation to the rest of the

world. In the initial equilibrium, the countries' policies are not co-ordinated. Let $\mathbf{P}^W$ denote the equilibrium of world prices and rest of the world is presented with a net supply function $-\mathbf{M}(\mathbf{P}^W)$. The two countries' net import vectors are $\mathbf{M}_A$ and $\mathbf{M}_B$, and $\mathbf{M}_A + \mathbf{M}_B = -\mathbf{M}(\mathbf{P}^W)$. The two countries then decide to co-ordinate their policies. Their agreed objective function is defined over the utilities of all consumers in the two countries, as is implied in the symmetry assumptions of perfect competition. This entails that all consumers and all producers face the same prices and that the optimum inter-personal distribution being achieved by lump-sum transfers. Thus, there is an agreed welfare function and this is maximised by permissible lump-sum transfers. In the initial equilibrium, producers and consumers in the two countries face different prices, as a result of consumption and/or production distortions. The two countries then co-ordinate policies by abolishing all tariffs between themselves set up a common external tariff and fully co-ordinate distributional policies between them. Let S^0 and S^1 represent the pre- and post-equilibrium situations wherein domestic taxes and subsidies are assumed to be non-existent:

$$(\mathbf{P}^{W0} - \mathbf{P}^{W1}) \cdot \mathbf{M}^0 + \mathbf{P}^{W1} \cdot \mathbf{T}^1(\mathbf{M}^1 - \mathbf{M}^0) \geq 0 \quad (2.23)$$

Where: $\mathbf{P}^W$ = vector of world prices; $\mathbf{T}$ = tariff vector; $\mathbf{M}$ = aggregate excess demand vector. If this is satisfied, then S^1 is preferable for both countries as a whole. The first term on the left-hand side indicates the terms of trade effect and the second the trade expansion (contraction) effect on the aggregate welfare of the two countries.

Dixit and Norman (1980) show that, to the extent that producer prices in the two countries differ in pre-equilibrium, they can increase their outputs of goods. The gains are larger the more different are the pre-equilibrium producer prices in the countries.

The literature emphasises the sufficient conditions for these trade reforms to be welfare-improving. The necessary conditions for welfare-improving reform however, are rarely considered. Restricting reform to those that are sufficient to ensure Pareto improvement may be too strict. A weaker requirement might be that no country necessarily loses from the reform.

Using the same procedure as before to find the sufficient conditions for a single small economy. There are n goods,⁴ with prices denoted by the column vector $\mathbf{P}$. As, before, $\mathbf{P}^1 = \mathbf{P}^{W1} + \mathbf{T}^1$. Assuming that tariff revenue is redistributed among consumers in a lump-sum fashion, the budget constraint for country j may be written as:

$$E_j(P_j, U_j) = R(P_j, V_j) + T_j M_j \quad (2.24)$$

⁴These could be final or intermediate goods. Intermediate goods enter the output vector as negative elements and pure intermediate goods enter the demand vector as zeros.

Where $E(\cdot)$ is the standard expenditure function of the economy, $R(\cdot)$ is the revenue function, U is the utility level and V_j is the factor endowments vector for economy j . Assuming all the standard properties, as before:

$$E'(\mathbf{P}, \mathbf{U}) = \mathbf{X}(\mathbf{P}, \mathbf{U}) \quad (2.25)$$

$$R'(\mathbf{P}, \mathbf{V}) = \mathbf{Y}(\mathbf{P}, \mathbf{V}) \quad (2.26)$$

$$Z(\cdot) = E'(\cdot) - R'(\cdot) \quad (2.27)$$

Where: (2.25) is demand, (2.26) is supply and (2.27) is the net trade vector. World prices – and from them, domestic prices – are determined by world market clearing conditions:

$$\sum_{j \in J} [E'_j(P_j, U_j) - R'_j(P_j, V_j)] = 0 \quad (2.28)$$

Where: J is the set of all countries. Thus, (2.25) and (2.28) can be used to solve for the endogenous variables in the system, namely equilibrium levels of utility and price. For any given country:

$$\begin{aligned} E(\mathbf{P}^0, \mathbf{U}^0) &= \mathbf{P}^0 \cdot \mathbf{X}(\mathbf{P}^0, \mathbf{U}^0) - \mathbf{P}^0 \cdot \mathbf{X}(\mathbf{P}^1, \mathbf{U}^1) + \mathbf{P}^0 \cdot \mathbf{X}(\mathbf{P}^1, \mathbf{U}^1) \\ &\geq \mathbf{P}^{0'} [\mathbf{X}(\mathbf{P}^0, \mathbf{U}^0) - \mathbf{X}(\mathbf{P}^1, \mathbf{U}^1)] + E(\mathbf{P}^0, \mathbf{U}^1) \end{aligned}$$

If $\mathbf{P}^0 \cdot [\mathbf{X}(\mathbf{P}^0, \mathbf{U}^0) - \mathbf{X}(\mathbf{P}^1, \mathbf{U}^1)] > 0$, then welfare must fall as a result of the change in tariff from $\mathbf{T}^0$ to $\mathbf{T}^1$. Thus, for welfare to rise:

$$\mathbf{P}^0 \cdot [\mathbf{X}(\mathbf{P}^0, \mathbf{U}^0) - \mathbf{X}(\mathbf{P}^1, \mathbf{U}^1)] \leq 0 \quad (2.29)$$

This says that a necessary condition for reform to be welfare-improving is that the new consumption bundle is not affordable at the old prices. Note that, if (2.29) is a necessary condition, so is:

$$\mathbf{P}^0 \cdot [\mathbf{X}(\mathbf{P}^0, \mathbf{U}^0) - \mathbf{X}(\mathbf{P}^1, \mathbf{U}^1)] - \mathbf{P}^0 \cdot [\mathbf{Y}(\mathbf{P}^0, \mathbf{V}^0) - \mathbf{Y}(\mathbf{P}^1, \mathbf{V}^1)] \leq 0 \quad (2.30)$$

is also necessary, since $\mathbf{P}^0 \cdot [\mathbf{Y}(\mathbf{P}^0, \mathbf{V}^0) - \mathbf{Y}(\mathbf{P}^1, \mathbf{V}^1)] \geq 0$ from profit maximising. Hence, an alternative form of the necessary condition for welfare to rise due to the reform is that:

$$\mathbf{P}^0 \cdot [\mathbf{M}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V}^0) - \mathbf{M}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}^1)] \leq 0 \quad (2.31)$$

In the small country case, world prices are given so that $\mathbf{P}^{W0} = \mathbf{P}^{W1}$ and assume the trade is balanced, thus:

$$\begin{aligned}\mathbf{P}^0 \cdot [\mathbf{M}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V}^0) - \mathbf{M}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}^1)] &= \mathbf{P}^{W0} \cdot [\mathbf{M}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V}^0) - \mathbf{M}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}^1)] \\ &\quad + \mathbf{T}^0 \cdot [\mathbf{M}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V}^0) - \mathbf{M}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}^1)] \\ &= \mathbf{T}^0 \cdot [\mathbf{M}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V}^0) - \mathbf{M}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}^1)]\end{aligned}$$

Hence, for a small economy to gain from the tariff reform, it is necessary for:

$$\mathbf{T}^0 \cdot [\mathbf{M}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V}^0) - \mathbf{M}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}^1)] \leq 0 \quad (2.32)$$

Thus, looking at the effects of reforms on consumption, import or tariff revenue evaluated at their pre-reform levels; if these fall short of pre-reform levels, welfare cannot be raised.

This can be summarised, as follows:

$$\mathbf{P}^0 \cdot [\mathbf{X}(\mathbf{P}^0, \mathbf{U}^0) - \mathbf{X}(\mathbf{P}^1, \mathbf{U}^1)] \leq 0 \quad (2.33)$$

or:

$$\mathbf{P}^0 \cdot [\mathbf{M}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V}^0) - \mathbf{M}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}^1)] \leq 0 \quad (2.34)$$

For a small economy with balanced trade, this can be written as:

$$\mathbf{T}^0 \cdot [\mathbf{M}(\mathbf{P}^0, \mathbf{U}^0, \mathbf{V}^0) - \mathbf{M}(\mathbf{P}^1, \mathbf{U}^1, \mathbf{V}^1)] \leq 0 \quad (2.35)$$

When applying these conditions to economic integration, as long as national welfare of the small economy increases, trade reform is welfare increasing. This is the same conclusion as that of Lipsey (1958).

2.3.5 Conclusions

Economic integration is a process of eliminating trade costs to increase countries' welfare. Traditional analysis of economic integration utilises a partial equilibrium approach, with Viner's development of the concepts of trade creation and trade diversion and emphasising that the welfare outcome is determined by the relative strength of these two effects. Meade's general equilibrium analysis shifts the focus to trade costs, in addition to trade creation and diversion. The theory of the 'second best' improves the theoretical fundamentals of the general equilibrium setting but also increases the ambiguity of any welfare analysis. By incorporating the 'transfer principle', the fundamentals are in place to facilitate the calculation of the effects of trade policy on welfare.

Allowing transfer payments between countries makes any trade bloc potentially favourable to all countries considering participating since they can be compensated for any losses resulting from membership. The analysis of trade blocs between countries can be extended to $n + 1$ countries and this implies that there is an incentive to extend a trade bloc until all countries are included; i.e., until global free trade prevails (Kemp and Wan 1976).

The final discussion establishes a welfare measure for comparing the welfare effects of trade policy changes.

2.4 Trade Costs, Economic Integration and Welfare

This section focuses on trade costs, their estimation and implications for welfare. Economic integration aims to reduce trade costs which, in turn, lead to greater harmonisation between segmented markets. Trade costs are an important variable in regard to a country's ability to participate in international trade and they have significant welfare implications because they are a hindrance to trade between countries. The main objective of this section is to review the theory of trade costs. It examines the definition of trade costs, transportation costs, tariffs and non-tariff measures. The welfare implications of trade costs are also examined.

2.4.1 Defining Trade Costs

Anderson and van Wincoop (2004) define trade costs broadly to include all costs incurred in getting a good to a final user other than the marginal cost of producing the good itself. This includes transportation costs (both freight and time costs), policy barriers (tariffs and non-tariff barriers), information costs, contract enforcement costs, costs associated with the use of different currencies, legal and regulatory costs and local distribution costs (wholesale and retail). Using this broad definition allows the measurement of trade costs to be identified as the divergence between the domestic price and the world price.

The direct evidence of trade costs comes in two major categories, those costs imposed by policy (tariffs, quotas and other trade barriers) and those by the environment (transportation, risk insurance and time). Direct transport costs include freight charges and insurance which is customarily added to the freight charge. Indirect transport costs include holding costs for goods in transit, inventory costs to buffer variability of delivery dates and preparation costs associated with shipment size (full containers versus partial loads). Indirect cost however, must be inferred. Wholesale and retail distribution costs enter retail prices in each country but, as these local trade costs are applied to both imported and domestic goods, they do not affect relative prices to buyers or the pattern of trade.

Bilateral trade costs are assumed to be a function of unobservables b_{ij} :

$$t_{ij} = f(b_{ij}) \quad (2.36)$$

The list of observable arguments, b_{ij} , used in the trade cost function in the literature includes directly measured trade costs, distance, adjacency, preferential trade membership, common language among others. The most common proxy for transport cost is distance; distance is most commonly assumed to have the following functional form:

$$t_{ij} = f(d_{ij}^p) \quad (2.37)$$

Where d_{ij} is distance between countries i and j .

Different measures are used to estimate distance in the literature but the most commonly used is the distance between capital cities. Where these differ from commercial centres, it is sometimes taken to be superior to use these distances although there can be difficulties if there is more than one commercial centre. The most reasonable measure is the bilateral distance between ports, supplemented by twice the land distance between ports and commercial centres.

In the case of preferential trade blocs, common languages and other such variables, implausibly strong regularity conditions are often implicitly imposed on the trade cost function. For example, the effect of membership of a preferential trade bloc on trade costs is often assumed to be uniform for all members. For custom unions, a uniform external tariff is indeed approximately the trade policy (although non-tariff barriers remain inherently discriminatory), while FTAs continue to have different national external tariffs and thus different effects.

International trade barriers can be decomposed into barriers associated with national borders and barriers associated with geographic frictions such as distance. National border barriers include policy barriers, in the form of tariffs and non-tariff barriers, languages, currencies, information and contracting costs and risk.

The model is completed by linking the unobservable trade costs to observables:

$$t_{ij} = d_{ij}^p b_{ij} \quad (2.38)$$

Where t_{ij} is the trade costs of bilateral trade flow between country i and country j , d_{ij} is the distance between country i and country j and b_{ij} is trade costs unrelated to distance or border costs.

2.4.2 *Transportation Costs*

The earliest attempt to introduce transport costs into the analytical models of trade is possibly that by Samuelson (1952) in the form of the 'iceberg' assumption. This is '... only a fraction of ice exported reaches its destination as unmelted ice

so will . . . a fraction of a country's exports . . . reach the other country as imports' (Samuelson 1954). The iceberg assumption is used extensively in analytical models. In addition to its simplicity, the assumption eliminates the need to treat transport as a separate sector producing a service and it avoids the complications that arise with different pricing of transport, e.g. f.o.b. and c.i.f. prices.

Krugman (1995) argues that the iceberg assumption ensures that the elasticity of demand is the same with respect to both the f.o.b. and c.i.f. price of a producing firm. Steininger (2001) points out that the biggest drawback of the assumption is that transport does not require any direct resource input. The amount of resources available to production within each country does not change and it is implicitly restricted to the same production technology as the transported good. It also implies that the production of transport is within the range of the factor intensity of the traded good. Falvey (1976) argues that the transport industry is capital-intensive and suggests the assumption that transport technology is identical across goods and countries; for any given factor price ratio, the same factor intensity is employed in transportation across goods and countries. The amount of transport service needed per unit transported may well differ across goods.

The export of goods and services involves time, effort and hence costs. Goods have to be physically loaded and unloaded, transported by truck, train, ship or plane, packed, insured etc. before they reach their destination. There they have to be unpacked, checked, assembled, and displayed before they can be sold to the consumer or an intermediate firm. All of these actions incur costs which tend to increase with distance. This can either be physical distance – which may be aggravated or alleviated by geographical phenomena such as mountain ranges or easy access to ports – or political, cultural and social distance – which also require time and effort. It is the very existence of transportation costs that supplies an incentive to locate production close to large markets. Anderson and van Wincoop (2004) define transportation costs as direct freight costs and indirect time costs. Transportation costs are a function of distance and time. Distance is directly related to freight charges and time is a function of distance, infrastructure, paperwork etc. Time costs are also a function of the institutional environment of a country. Limao and Venables (2001) emphasise the dependence of transportation costs on infrastructure, measured as an average of the density of the road network, the paved road network, the rail network and the number of telephone main lines per person.

2.4.3 *Tariff Barriers*

A tariff can be defined simply as an indirect tax on imports. The most usual forms of a tariff are *ad valorem* – where a specific percentage of the value is applied to an imported good, and/or *specific* – where a specified amount is applied to each unit of an imported good. *Ad valorem* tariffs have the feature of being index linked while

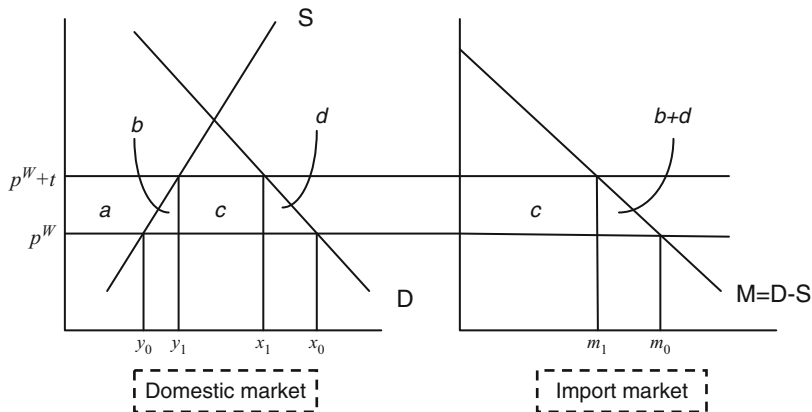


Fig. 2.1 Implications of a tariff for a small country

specific tariffs reduce the incentive for under-invoicing and other illegal practices. These two tariff types are often combined with the total tariff equal to a specific amount plus a percentage of the price. There are several other forms of tariff. Progressive tariffs permit a specific quantity or value of a good to be imported at the ‘normal’ tariff rate while a higher rate is charged on those imports that exceed the specific quantity or value. A seasonal tariff is a special case where a different tariff rate is applied to an imported good, depending upon the time of year.

Tariffs are the only form of trade protection permitted by the WTO rules. Exceptions to MFN are permitted only under specific circumstances, including: the ‘escape clause’ under which countries may increase tariffs temporarily as a result of injury to an import-competing industry (GATT Article XIX); and anti-dumping duties under which tariffs are applied to offset import prices that are deemed ‘too low’ (GATT Article VI).

Using the same notation as in previous sections, the change in welfare as a result of the imposition of a tariff is:

$$\Delta W = \mathbf{T}^1 \cdot (\mathbf{M}^1 - \mathbf{M}^0) - (\mathbf{P}^{W0} - \mathbf{P}^{W1}) \cdot \mathbf{M}^0 + \Pi \quad (2.39)$$

Where, Π are the profits from the import-competing industry, $\mathbf{P} \cdot \mathbf{Y} - C(\mathbf{Y})$. Under perfect competition, $\Pi = 0$, and for a small country the change in welfare is:

$$\Delta W = \mathbf{T}^1 \cdot (\mathbf{M}^1 - \mathbf{M}^0) \quad (2.40)$$

For a tariff to be welfare improving, $\mathbf{T}^1 \cdot (\mathbf{M}^1 - \mathbf{M}^0)$ needs to be ≥ 0 . The critical point is at a zero tariff.

From Fig. 2.1, before the imposition of a tariff, domestic demand is x_0 and supply is at y_0 , so imports are $m_0 = x_0 - y_0$. When a tariff is imposed on an imported good,

the equilibrium domestic price changes by the amount of the tariff, to $p = p^W + t$. This leads to reduced demand of x_1 and increased supply of y_1 . The change in welfare is $-(a + b + c + d)$. Consumer surplus loss is a , producer surplus gain is c and tariff revenue is $-(b + d)$, which is always negative. Area $(b + d)$ is the deadweight loss of the tariff. The area d is interpreted as the consumer surplus loss for those units no longer purchased ($x_0 - x_1$), while area b is interpreted as the increase in marginal cost for the extra units produced ($y_1 - y_0$).

The deadweight loss is measured by the triangle under the import demand curve and can be measured as the fraction of import expenditure:

$$\frac{\Delta W}{p \cdot m} = \frac{\mathbf{T}^1 \cdot (\mathbf{M}^1 - \mathbf{M}^0)}{\mathbf{P} \cdot \mathbf{M}^0} \quad (2.41)$$

The critical point at $\mathbf{T} = \mathbf{0}$ is a global maximum. The optimal tariff for a small country is zero. The deadweight loss may well be larger due to less import variety when a tariff is imposed (Feenstra 1992; Romer 1994).

The result that a tariff is harmful for a small open economy relies on the assumption that there are no distortions in the economy. If there are distortions, tariffs could be used to offset these distortions and thereby increase welfare. This possibility is an application of the theory of second best, which states that in the presence of multiple distortions, welfare is not necessarily improved by removing a single distortion. An equivalent statement is that, in the presence of distortions, adding an additional distortion may improve welfare.

Using the same tariff analysis for the case of a large country:

$$\Delta W = \mathbf{T}^1 \cdot (\mathbf{M}^1 - \mathbf{M}^0) - (\mathbf{P}^{W_0} - \mathbf{P}^{W_1}) \cdot \mathbf{M}^0 \quad (2.42)$$

From Fig. 2.2, the change in welfare is $-(a + b + c + d)$, consumer surplus loss a , producer surplus gain $(c + e)$ and tariff revenue is $e - (b + d)$. The area e is terms of trade gain, $-(\mathbf{P}^{W_0} - \mathbf{P}^{W_1}) \cdot \mathbf{M}^0$. The optimal tariff is:

$$\mathbf{T}^1 \cdot \mathbf{M}^1 - \mathbf{T}^1 \cdot \mathbf{M}^0 = \mathbf{P}^{W_0} \mathbf{M}^0 - \mathbf{P}^{W_1} \mathbf{M}^0 \quad (2.43)$$

$$\mathbf{T}^1 = \frac{\mathbf{P}^{W_0} \mathbf{M}^0 - \mathbf{P}^{W_1} \mathbf{M}^0}{\mathbf{M}^1 - \mathbf{M}^0} \quad (2.44)$$

Which depends on the slope of the foreign export supply curve, $\frac{\Delta \mathbf{P}^W}{\Delta \mathbf{M}}$.

The optimal percentage tariff equals the inverse of the elasticity of the foreign export supply. This complies with the terms of trade argument: a country that is large enough to influence the prices at which it trades can increase its level of welfare relative to the free trade benchmark by restricting trade below the free trade level. The rationale for government intervention is a terms of trade externality because individual agents cannot by themselves exploit their joint market power. Hence, the government must harness the country's market power through a tax on

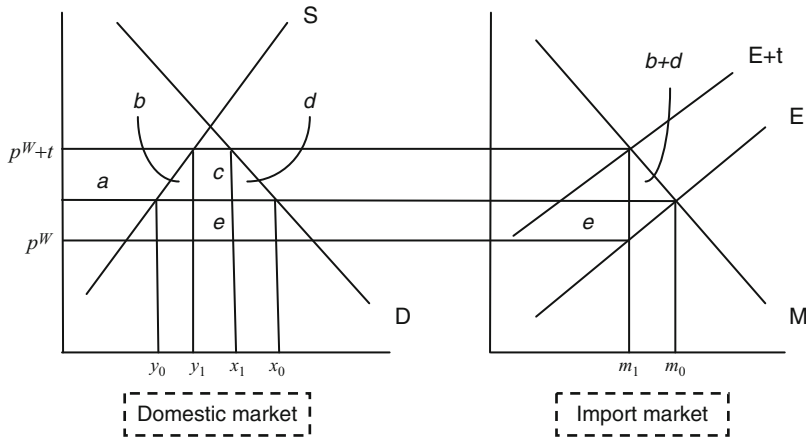


Fig. 2.2 Tariff implications for a large country

trade. In the case of a large importing country, the optimal import tariff in *ad valorem* terms is equal to the inverse of the foreign export supply elasticity.

2.4.4 Non-Tariff Barriers

The preceding discussion assumes that trade barriers take the form of *ad valorem* import tariffs. With minor modifications, the approach is also applicable to non-tariff barriers (NTBs). Like tariffs, NTBs drive a wedge between the price of a product in the supplier's domestic market and the price faced by consumers in an importing partner country.

Reinterpreting t as the *ad valorem* rate of costs imposed by NTBs to imports of a commodity, Figs. 2.1 and 2.2 also illustrate the effects of a unilateral non-discriminatory imposition of the NTBs faced by foreign suppliers of the importable good. The domestic price, production, consumption and trade effects are the same as in the case of the imposition of an import tariff at rate t . Therefore, one may refer to the NTB cost mark-up t as the tariff equivalent of NTBs in the present context. While the importance of tariffs has been steadily reduced, the relative importance of NTBs has increased.

2.4.5 Trade Costs and Economic Welfare

Trade costs include the imposition of trade taxes (i.e. specific departures from free trade) and/or costs that result in, $P \neq P^W$. As governments have direct control over trade taxes, such departures can be rationalised on four grounds:

1. Trade taxes can be used to affect world market prices and thereby to achieve improved terms of trade.
2. Trade taxes can be used to achieve an optimum domestic distribution of income.
3. Trade taxes can be used to achieve exogenously given targets for trade, production or domestic consumption.
4. Trade taxes can be used to correct distortions resulting from market failure.

Nearly all countries have made use of tariffs and other types of trade policies, especially in the early stages of their development.

In the case of a small country, the excess supply function is perfectly elastic, so world prices are fixed, as before. The economy goes from free trade to the imposition of trade taxes. Under free trade $\mathbf{T} = \mathbf{0}$ and $\mathbf{P} = \mathbf{P}^W$. With the imposition of trade taxes, $\mathbf{T} \neq \mathbf{0}$ and $\mathbf{P}^W + \mathbf{T} = \mathbf{P}$. The welfare measure is:

$$\Delta W = \mathbf{T} \cdot (\mathbf{M}^1 - \mathbf{M}^0) - (\mathbf{P}^W - \mathbf{P}^{W1}) \cdot \mathbf{M}^0 \quad (2.45)$$

$$\mathbf{P}^W - \mathbf{P}^{W1} = \mathbf{T} \text{ and } \mathbf{M} = \mathbf{X} - \mathbf{Y}$$

$$\begin{aligned} \Delta W &= (\mathbf{P}^W - \mathbf{P}^{W1}) \cdot (\mathbf{M}^1 - \mathbf{M}^0) \\ &= \mathbf{T} \cdot (\mathbf{M}^1 - \mathbf{M}^0) \end{aligned} \quad (2.46)$$

So, welfare increases if and only if import increases.

2.4.6 Welfare Comparison with Consumer Distortions

Trade costs generate distortions in an economy, which can be direct as in the case of consumption distortions, as a result of prices not being equalised across countries, and indirect as in the case of production distortions, where trade barriers for example result in inefficient production. The first fundamental welfare theorem establishes the perfectly competitive case as a benchmark for thinking about outcomes in market economies. In particular, any distortions that arise in a market economy – and hence any role for Pareto-improving market intervention – must be traceable to a violation of at least one of the assumptions of this theorem. If these distortions are absent, the atomistic consumer and firms will jointly choose $\mathbf{X}$ and $\mathbf{Y}$. But consumption distortions may also exist in the economy.

Following Grinols and Wong (1991), consumption distortions are constraints on consumption choices other than the budget constraint. These constraints can be parameterised by a variable α , which may be a scalar or vector. The expenditure function in the presence of constraint α can be denoted as $E(\mathbf{P}, \mathbf{U}; \alpha)$ and is no less than the expenditure function without the constraint; $E(\mathbf{P}, \mathbf{U}) \leq E(\mathbf{P}, \mathbf{U}; \alpha)$. The quantities of traded goods and factor services can be denoted by the vector $\mathbf{M} = \mathbf{X} - \mathbf{Y}$ – where the import components of $\mathbf{M}$ are positive and the export components

of $\mathbf{M}$ are negative. When taxes are imposed on traded goods and factor services, divergences between the domestic price $\mathbf{P}$ and international price $\mathbf{P}^W$ are created, with the tax revenue being equal to $(\mathbf{P} - \mathbf{P}^W) \cdot \mathbf{M}$. The trade balance requirement implies that $\mathbf{P}^W \cdot \mathbf{M} - \mathbf{A} = 0$, where $\mathbf{A}$ represents borrowing from foreigners to meet the deficit or lending to foreigners if there is a surplus. The welfare measure with distortion is:

$$\Delta W = E(\mathbf{P}^1, \mathbf{U}^1; \alpha^1) - E(\mathbf{P}^1, \mathbf{U}^0; \alpha^1) \quad (2.47)$$

Using the definition of the expenditure function:

$$\begin{aligned} E(\mathbf{P}^1, \mathbf{U}^1; \alpha^1) &= \mathbf{P}^1 \cdot \mathbf{X}^1 \\ &= \mathbf{P}^1 \cdot \mathbf{Y}^1 + \underbrace{(\mathbf{P}^1 - \mathbf{P}^{W1}) \cdot \mathbf{M}^1}_{\text{tariff revenue}} + \mathbf{A}^1 \end{aligned} \quad (2.48)$$

where the external balance condition has been used. Using (2.47) and (2.48) and rearranging the terms, the measure of welfare change can be given alternatively as:

$$\Delta W = (\mathbf{P}^1 - \mathbf{P}^{W1}) \cdot (\mathbf{M}^1 - \mathbf{M}^0) + (\mathbf{A}^1 - \mathbf{A}^0) + (\mathbf{P}^{W1} - \mathbf{P}^{W0}) \cdot \mathbf{M}^0 + \eta_c \quad (2.49)$$

Where the consumption effect is:

$$\eta_c = \mathbf{P}^1 \cdot \mathbf{X}^0 - E(\mathbf{P}^1, \mathbf{U}^0; \alpha^1) \quad (2.50)$$

The measure of welfare change can be disaggregated into five terms. Firstly, the revenue effect ($\mathbf{R}$), which is the difference between the domestic price and world price at S^1 , times the difference between imports at time $\{0, 1\}$. If $\mathbf{T}^1 = 0$, then $\mathbf{P}^1 = \mathbf{P}^{W1}$. This effect on welfare is the revenue effect ($\mathbf{R}$). Secondly, the change in transfers to or from foreigners through lending or borrowing is $(\mathbf{A}^1 - \mathbf{A}^0)$, which is the transfer effect ($\mathbf{F}$). Thirdly, the terms of trade effect which is the change in world price $\mathbf{P}^W$ before and after, which is zero for a small country. The change in welfare is the sum of these effects. The revenue effect $\mathbf{R} = (\mathbf{P}^1 - \mathbf{P}^{W1}) \cdot (\mathbf{M}^1 - \mathbf{M}^0)$, is a measure of the increase in tariff revenue, calculated using the final tariff rates, when the levels of imports increase from $\mathbf{M}^0$ to $\mathbf{M}^1$. If there is free trade or if the level of imports does not change, the revenue effect is zero. The transfer effect, $\mathbf{F}$, measures the welfare effects of a change in the international transfer. The consumption effect, η_c , measures the difference between the expenditure to purchase the initial consumption bundle and the minimum expenditure needed to reach the initial utility level in the final situation.

Equation 2.49 is applied to calculate the change in welfare of the economy when moving from one situation to another. For a small country the change in welfare is:

$$\Delta W = (\mathbf{P}^1 - \mathbf{P}^{W1}) \cdot (\mathbf{M}^1 - \mathbf{M}^0) + (\mathbf{A}^1 - \mathbf{A}^0) + \eta_c \quad (2.51)$$

If the constraints in consumption remain the same or are absent in these two situations, then $\alpha^0 = \alpha^1$. By definition given the expenditure function, the consumption effect is always non-negative; that is $\eta_c \geq 0$. Then $\mathbf{P} \cdot \mathbf{Y}^1 \geq \mathbf{P} \cdot \mathbf{Y}^0$ (this is the case if production is perfectly competitive). In general, because of the presence of distortions, the consumption effects can be negative or zero. The sign and magnitude depend on substitution between goods in consumption, prices and the constraints in the economy. If substitution in consumption is ruled out, the consumption effect is zero.

2.5 The Impact of Size on the Characteristics of Economies

There is no definitive measure of size in the economic literature apart from the general view that a country is small if it cannot affect world prices for its imports and exports. This is the terms of trade of the small economy definition of Johnson (1960). Marcy (1960) argues that a nation is small from the point of view of foreign trade when its dependence on foreign markets is relatively great but its contribution to them small in absolute terms. Very few countries, if any, in the global economy according to Armstrong and Read (1998) possess the power to influence their own terms of trade such that virtually all countries are ‘small’ in economic terms. Magee and Magee (2008) conclude that the US is a small country in world trade using this measure of size since their trade policies have negligible impacts on world prices. This section considers the meaning of economic size and identifies the key characteristics of small economies. It then discusses the issues of trade costs and the impact of economic integration in the context of these arguments.

2.5.1 *The Analysis of the Size of Economies*

There is no general agreement in applied economic analysis on a numerical measure of country size. Early examples of the literature (e.g. Marshall 1919; Ohlin 1933) use geographical measures such as land area or arable land as the measure of size. In the middle of last century, several economists turned to using demographic measures, notably population, as a measure of size (Kuznets 1960; Marcy 1960; Michealy 1962). Population remains the most common measure of size and is used by all of the principal international organisations, such as the UN and World Bank. The primary reason for this is ease of availability of population data as well as the fact that it provides a crude proxy for the size of both the domestic market and the local labour force (Armstrong and Read 1998).

From an economic viewpoint however, these measures have drawbacks in that there needs to be a measure of the size of a market which incorporates value. A country’s market size is its domestic market together with its participation in external trade, determined by the extent of its openness to international trade. In a closed economy with no external trade, the domestic market is the same as a country’s market size.

To measure the market size of a country, a measure is needed of the money value of the goods and services generated by domestic economic activity.

The measurement best suited for this purpose is GDP (Young 1928). GDP measures the market size of a particular region as its domestic market and part of the world market and is therefore an ideal economic measure of size. When looking at size of an economy the important thing is the comparison; a country may be small compared to one country but large compared to another. A comparison of GDP between countries or group of countries enables the classification of small and large.

Balassa (1962) identifies several issues with respect to the size of economies. First, the population measure does not reveal the possibilities of the division of labour and might intimate (falsely) that, for instance, the economic size of China or India is greater than the United States. Geographical measures as an indicator of economic size are best judged by comparison; Greenland is six times larger in area than Germany and Iceland more than twice as large as Denmark. Balassa considers the volume of production to be the most appropriate yardstick. According to this measure, the size of a national economy or trade bloc is given by its GNP. This view supports Edwards (1960) argument that the size of a nation's economy is roughly reflected by the size of its GNP. Further qualifications are required if internal differences in tastes and transportation costs are considered. For a given GNP, the greater the costs of transportation and the more diverse are tastes, the smaller is the effective size of the market.

From an economic perspective, the size of a market should reveal the minimum efficient scale (MES) of output that can be undertaken within its borders (Armstrong and Read 1998). Differences in the sizes of nation can be regarded primarily from the point of view of the opportunities for attaining the greatest possible economies and regularity of production, in so far as these depend on scale of production (Robinson 1960). Scitovsky (1960) argues that, technologically an economy is too small if its market is too small to provide adequate outlets for full-capacity output of its most efficient productive plants in any given industry. The minimum efficient size of an economy is generally different for different industries. An economy can be said to be too small if it fails to provide the competitive conditions necessary to spur utmost efficiency and lead to the establishment of the technically most efficient plants. The technological optimum is probably reached very much sooner than the economic optimum. This is the sub-optimality argument regarding the size of economies (see Armstrong and Read 1998). This encompasses the impact of small market size on economies of scale, indivisibilities, efficiency and competitiveness (Scitovsky 1960) and economies of scope economies (Streeten 1993). The degree of sub-optimality is a function of market size and technology, which makes the definition of economic size rather elusive as a measure but technically correct.

2.5.2 The Particular Characteristics of Small Economies

Several studies together identify the specific economic characteristics of small countries (Kuznets 1960; Robinson 1960; Michealy 1962; Lloyd 1968; Khalaf 1971; Jalan 1982) as:

- A limited ability to achieve economies of scale.
- Greater concentration of domestic production.
- Greater concentration of exports.
- Greater openness to foreign trade.
- Greater geographical concentration of foreign trade.
- A relatively large public sector.

From the viewpoint of the theory of economic integration theory, the most important characteristic of a small country is the scale economy constraint. This source of sub-optimality means that, production cannot reach the MES in a wide range of activities given a small market. In addition domestic competition is likely to be limited given the small number of feasible firms in many industries (Armstrong et al. 1993; Armstrong and Read 1998). The implications are that sub-optimal size is associated with economic disadvantage (Armstrong and Read 1995, 1998; Briguglio 1998):

- Limited possibilities for economies of scale.
- Limited possibilities for the development of endogenous technology.
- Limited natural resource endowments and high import content in domestic output.
- A dependence of export markets
- Problems of public administration
- High transport and co-ordination costs (remoteness).

It is easier for monopolies to develop given the smaller number of firms of minimum efficient size that can be supported by a small domestic market. Further, competition is more likely to be oligopolistic and their reactions are likely to be aggressive rather than co-operative (Robinson 1960). Kuznets (1960) concludes that the economic structure of small nations is typically less diversified than that of large nations, and specifies three reasons. First, the impact of the area and its limited effect on the supply of natural irreproducible resources. Second, the conflict between the MES of plant for some industries and the limited domestic market. Third, while a small nation may lack many natural resources, its supply of a few may give it a marked comparative advantage over larger countries whose supply may be as large or larger absolutely, but much lower on a per head or per unit of need basis – notably in the oil-rich small economies of the Middle East. If a small nation has even a relatively small supply of some resource of world-wide use, it may well concentrate on this valuable resource to the point where little labour force and few other resources are left for other domestic production, excepting, of course, essential goods that cannot be imported.

2.5.2.1 Openness to International Trade

The smaller the country, the larger the likely ratio of exports and imports to total output (Kuznets 1960). The critical importance of tradeable goods to small

economies thus necessitates the pursuit of highly open trade regimes with limited scope for import-substitution and infant industry protection policies (Armstrong and Read 1998). Openness is a function of trade costs, so trade barriers in small countries can be expected to be relatively low. Small countries face incentives to adopt open trade policies because they cannot benefit from access to larger markets unless they are open and internationally competitive. Thus, small countries can be expected to be more open to trade (Kuznets 1960; Alesina and Wacziarg 1998; Armstrong and Read 1998). External trade can and does provide an effective escape from the disadvantages of smallness and, in general, the smaller the economy the higher the ratio of imports to GDP.

2.5.2.2 Exports

International trade has a greater weight in the economic activity of small nations than that of larger ones. This is particularly true of nations that have developed and attained fairly high levels of per capita output and consumption (Kuznets 1960; Armstrong et al. 1998). The export sector in small countries therefore has greater importance in output and exports per capita can be expected to be substantially higher than in larger economies. Armstrong and Read (1998) point out that the structure of the export sector in small economies is highly specialised and undiversified due to the constraint on feasible domestic activities. This gives rise to over-dependence on one or two export products and export markets (Briguglio 1995; Armstrong and Read 2006).

In the case of industrial countries with small geographic areas, the dependence upon external markets is due to scarcity of natural resources and limited population. These countries are thereby prevented from diversifying their economy and hence must rely heavily on imports. Since their domestic markets provide insufficient outlets for production, they must export a large part of it (Marcy 1960).

2.5.2.3 Export Concentration and Trading Partners

Accompanying the heavier reliance of small countries on foreign trade is a greater tendency to rely on imports from and exports to a limited number of countries – with such concentration particularly noticeable in exports (Kuznets 1960). Armstrong and Read (1998) argue that this geographic export concentration exacerbates the problems of output and export concentration. Export diversification is smaller in smaller countries and distance to trading partners closer. A nation is small when it has a high degree of dependence upon foreign trade and few markets for its exports. In such circumstances, it must inevitably suffer a domination effect on the part of its buyers (Marcy 1960).

2.5.2.4 Import Variety

Consumers not only demand a good, they also demand varieties of the good. They prefer more varieties of a single good to fewer. The number of varieties for a single good should be closely related to the size and income level of the market. Where larger size of the market should result in a larger variety for each single good and higher income level should result in an increased number of varieties for each single good. A small market, however wealthy, may therefore be unable to sustain a full range of varieties.

2.5.2.5 Trade Vulnerability

The term vulnerability refers to proneness to damage from external forces (Briguglio 1998). A critical feature of small countries is their high level of structural openness to international trade which greatly impacts their vulnerability of small countries because of their extremely sensitivity to developments in the global trading environment. Small countries can therefore be expected to be disproportionately affected by both increasing trade liberalization and the growth of regional trade blocs (Armstrong and Read 1998). The effects of changes in trade policies can be expected to affect the economies of small countries more than large countries because of their greater dependence on international trade. The openness of small countries may exacerbate the inherent vulnerability by intensifying their exposure to exogenous shocks.

International trade vulnerability refers to permanent or quasi-permanent international trade conditions which expose a country to external risk factors and is a function of degree of openness, export product concentration and export market diversification. The higher the openness to trade, the greater the vulnerability and the lower the export product concentration and export market diversification is, the more is the vulnerability. The greater the risk (trade vulnerability), the higher the trade performance in good years and poorer in bad years. It can be argued however, that trade openness reduces vulnerability since those small countries with the highest incomes are also most open to trade (Armstrong and Read 2002). Nevertheless, external shocks are exaggerated in a small economy due to its greater exposure to external factors.

2.5.2.6 The Public Sector

The diseconomies inherent in small countries also apply to their ability to provide effective administration and governance (Selwyn 1975). These diseconomies arise from indivisibilities in public administration. Smaller countries cannot share the cost of partially or completely non-rival public goods over large populations, so per capita expenditure on these goods is higher in small countries than in large ones.

Government consumption as a share of GDP is therefore likely to be smaller in larger countries (Kuznets 1960; Alesina and Spolaore 2003). For instance, there are fixed costs in establishing a set of institutions and legislative power. Kuznets (1960) however, also argues that many small countries eschew substantial defence expenditure, so partially redressing their need for disproportionate spending by taking free-riding on the defence umbrellas of larger countries (Armstrong and Read 2002).

2.5.2.7 Economic Performance

Scitovsky (1960) concludes that the disadvantages of economic smallness all relate to competition. Competition encourages economic efficiency and progress if the economic unit is large enough. These effects are likely to be achieved by freer trade and economic union alike, provided that they promote economic, social, and intellectual contact to a significant extent. Freer trade or economic union promotes economic and technical efficiency by rendering competition less personal and thereby more effective. Relations among small producers comprising an industry are often so close and friendly as to keep each of them from engaging in competitive actions that would hurt the rest.

Krugman (1994) criticises the use of the concept of competitiveness in the context of economic performance, arguing that it renders it meaningless.

... it is simply not the case that the world's leading nations are to any important degree in economic competition with each other, or that any of their major economic problems can be attributed to failures to compete on world markets ... Competitiveness is a meaningless word when applied to national economies (Krugman 1994, pp. 30, 44).

The problem with Krugman's argument is that he uses comparative advantage to draw conclusions. Ricardo's theory explains a country's comparative trading advantage in terms of international differences in productivity rather than the competitiveness of countries in particular products. It neglects differences in product market conditions. Linder (1961) argues that nations tend to export differentiated goods for which there exists domestic demand. Vernon (1966) supports the argument that domestic market conditions may provide firms with technological advantages that can be utilised through exports to provide a competitive edge. Porter (1998) argues that domestic competition can create pressure for improvements through innovations in ways that upgrade the competitive advantages of nations. Domestic competition in product markets is closely related to trade performance, which should lead to greater penetration of domestic products into foreign markets. In turn, this should translate into greater export market diversification, an increased number of trading partners and increased exports per capita.

In spite of the disadvantages facing small nations, many of them, have shown very strong economic performance (Armstrong and Read 1995, 1998; Armstrong et al. 1998; Milner and Weyman-Jones 2003). Armstrong and Read (1998) explain the success of small states through their openness and export-led growth, which

simultaneously reduce the impact of their scale economies constraint and forces firms to be internationally competitive. Other possible factors that have contributed to their growth success are their small size such that they have been able to negotiate favourable PTAs with major trading partners (Armstrong and Read 1995, 2002; Armstrong et al. 1998). Further, many small states have been able to exploit their cultural distinctiveness and strong social capital (Baldacchino and Milne 2000; Armstrong and Read 2002; Baldacchino 2004). Some small states also possess valuable natural resources (e.g. fish, oil) while others have benefited from being located in affluent regions of the world (Armstrong et al. 1998; Armstrong and Read 1998, 2002). The proximity of small states to high growth countries and regions may have additional beneficial economic performance effects, notably convergence 'club' effects in Europe and the Caribbean (Armstrong et al. 1998).

Successful economic performance in small nations is likely to be founded upon niche markets and human-capital intensive activities, such as tourism and financial services (Armstrong and Read 1995; Armstrong et al. 1998). Economic performance is closely linked to adaptability. Although small countries face greater risks in their export markets than larger countries, their flexibility in adapting to changes in market conditions is critically important for their survival. Factors that affect their adaptability are human capital, social cohesion and technological awareness – a function of domestic human capital and openness to trade. Competition is also important in supporting adaptability. The type, quality and cost of infrastructure, e.g. transportation, communications and banking systems, affects competition.

2.5.3 Trade Costs and the Size of Economies

An interesting perspective is to examine the relationship between the size of economies and trade costs since the latter are a function of openness. This theoretical discussion draws on work by several authors: Ohyama (1972), Alesina et al. (2000, 2005), Alesina and Spolaore (2003), Anderson and van Wincoop (2004) and Spolaore and Wacziarg (2005).

2.5.3.1 The Size of Economies and Productivity

In each country j , the Cobb-Douglas production function is:

$$Y_j = A_j \cdot K_j^\alpha \cdot L_j^{1-\alpha} \quad ; \quad j = 1, \dots, k \quad (2.52)$$

Labour and capital are immobile between countries. A_j is the total factor productivity and depends upon total accumulated education denoted as H_j . This is a function of the size of the economy and its international relations. In autarky:

$$A_j = H_j \quad (2.53)$$

Which shows the importance of size. With complete international economic integration, i.e. no trade barriers, everyone can learn from everyone else independent of location, for every country j :

$$A_j = \sum_{j=1}^k H_j \quad (2.54)$$

With no trade barriers the importance of size is eliminated. In a world of trade barriers, the size of a country's market is influenced by its political borders. With free trade, the size of countries is irrelevant for the size of markets, so the size of a country is unrelated to its productivity (Hobsbawn 1990; Wittman 1991).

2.5.3.2 The Size of Economies and International Trade

In each country j , a specific intermediate input Q_j is produced using the country-specific capital according to a linear production function:

$$Q_j = K_j \quad (2.55)$$

Intermediate inputs can be traded internationally in perfectly competitive markets. The Cobb-Douglas production function takes the form:

$$Y_j = A_j \cdot Q_j^\alpha \cdot L_j^{1-\alpha} \quad (2.56)$$

Labour and capital are immobile between countries. When the intermediate good is traded internationally, a trade cost is incurred. Using the standard 'iceberg' assumption to model trade costs (Sect. 2.4.2); if one unit of an intermediate good produced in country j is exported to country k ; only $(1 - t_{jk})$ units of the intermediate good arrive. The parameter t measures trade costs. Let D_j denote the units of intermediate inputs used domestically and F_{jk} denote the units of good shipped to location k . Only $(1 - t_{jk})F_{jk}$ units will reach the final user. In equilibrium, markets are perfectly competitive ($P = MP$ domestically and internationally). Let $A_j = A$:

$$P_j = \alpha A D_j^{\alpha-1} = \alpha A (1 - t_{jk})^\alpha F_{jk}^{\alpha-1} \quad (2.57)$$

Where P_j is the price of input in country j . Defining the size of countries as, $s_j = Y_j / \sum Y_j$, this definition shows size as percentage of world output, i.e. compared to other economies. From, $Q_j = K_j$, it follows that resource constraint for each input is:

$$s_j D_j + \sum s_j F_{jk} = K_j \quad (2.58)$$

where K_j is the stock of capital in country j . By substituting (2.57) into (2.58):

$$D_j = \frac{K_j}{s_j + \sum s_j (1 - t_{jk})^{\alpha/(1-\alpha)}} \quad (2.59)$$

and:

$$F_{jk} = \frac{(1 - t_{jk})^{\alpha/(1-\alpha)} K_j}{s_j + \sum s_j (1 - t_{jk})^{\alpha/(1-\alpha)}} \quad (2.60)$$

Anderson and van Wincoop (2004) assume that trade cost is $t_{jk} = t_{kj}$. For now, this assumption will be kept for simplicity, and define:

$$\omega = (1 - t)^{\alpha/(1-\alpha)} \quad (2.61)$$

The lower the trade cost, the higher the ω . Demand is unbounded. Equations 2.59, 2.60 and 2.61 imply that:

$$D_j = \frac{K_j}{s_j + (1 - s_j)\omega} \quad (2.62)$$

Where D_j is the domestic demand in country j , and:

$$F_{jk} = \frac{\omega K_{jk}}{s_j + (1 - s_j)\omega} \quad (2.63)$$

Where F_{jk} is the foreign demand for the intermediate good. Households in each country maximize their utility, $U(X)$. In equilibrium $Y = X$:

$$Y = X = A \cdot K^\alpha [s_j + (1 - s_j)\omega]^{1-\alpha} \quad (2.64)$$

Output and consumption are increasing in ω , increasing in country size s_j and decreasing in $s_j \cdot \omega$ (country size multiplied by trade costs). The effect of s_j is smaller the larger is ω :

$$\frac{\partial Y_j}{\partial \omega} \succ 0; \quad \frac{\partial Y_j}{\partial s_j} \succ 0; \quad \frac{\partial^2 Y_j}{\partial s_j \partial \omega} \prec 0 \quad (2.65)$$

These results show that the economic benefits from size are decreasing in ‘openness’ – lower trade costs – and the economic benefits from ‘openness’ are decreasing in size. This result implies that large countries have higher average trade costs than small countries.

2.5.4 *Economic Integration and the Size of Economies*

The practical argument for economic integration agreements is based on the premise that there are economies of scale which are not exhausted within the limits of the size of nations but which could be achieved within the limits of a larger trading area (Robinson 1960). Marcy (1960) emphasises the effect on competition and argues that the elimination of custom duties, quotas and government subsidies puts all competitors on an equal footing. Any differences are due only to the distance between the location of production and the point of sale. Initially, it is quite possible that the elimination of protective measures is felt much more painfully by the firms of the formerly most highly protective member countries. Firms which sell highly specialised quality products with inelastic demand will be in the most favourable position.

Neo-classical trade theory predicts that large countries tend to be net exporters in scale-intensive industries (Helpman and Krugman 1985; Krugman and Venables 1995). Large countries should therefore gain more from economic integration, especially in sectors where there exists scale economies that can be exploited in an even larger market, as a result of lower trade costs. Other things being equal, if two countries, one large and one small, have the same *ad valorem* trade costs, the larger country can be expected to gain more from economic integration because of economies of scale. Casella (1996) however, argues that enlarging a trade bloc increases the size of the market to which all countries have access; this increase is more significant for firms located in small countries, whose own domestic market is small. The increase in competitiveness is therefore relatively larger for small countries, so the entry of new members in a trade bloc will particularly favour small countries and guarantees their access to the market of larger partners (Graham 1923; Balassa 1962, 1967).

Marcy (1960) argues that small nations are obliged to sell abroad a considerable proportion of their production and find it very hard to do so because of hindrances caused by trade barriers, production subsidies and foreign exchange controls. Because their domestic markets are limited, they must limit their productive capacity rather than increase their productivity and reduce costs. Marcy concludes that small countries should be integrated into a larger group within which there is full mobility of goods, services and, if possible, labour and capital. Small nations would have thereby gained access to very large markets and realise scale economies. Scitovsky (1960) argues that, if an economy is too small technologically – in providing insufficient market outlets for output, even of a single modern and efficient plant – then international trade is of little avail. This is because for mass-production methods to be profitable require market outlets that are large, homogeneous and stable over time. These are not the usual characteristics of an export market. In this case, economic integration is preferable, provided that it guarantees not only free and unrestricted trade but also complete stability of exchange rates among members.

Underlying the conclusion of all this literature however, is the assumption that economic integration widens the extent of the market of small countries and enables

them to achieve economies of scale that were previously inaccessible because of their size. It has been pointed out however, that the critical imperative of a high degree of openness to trade in small countries by implication negates the perceived gains from membership of a trade bloc (Rothschild 1944, 1963; Read 2004). Hirschman's (1958) core-periphery effects also suggest that scale-extensive production activities tend to agglomerate close to large population centres, i.e. in the larger more central member states of a regional economic bloc, so reducing the possible gains for small countries from membership of a larger market. Membership of a trade bloc is also likely to diminish the trade policy autonomy of small members, which is likely to have adverse implications for their welfare (Read 2004). Read concludes that the general assumption that small nations benefit from membership of regional economic blocs is by no mean as clear-cut as the general literature assumes but rather that the critical issue is the extent to which such economic integration agreements generate efficiency gains and reduce uncertainty through closer trade links for their member states.

2.5.5 *Welfare Comparisons with Production Distortions*

The sub-optimality argument can be viewed as a production distortion and, as such, can be classified into two types. Firstly, firm-level constraints; such that, given prices, the aggregate output chosen jointly by firms, $\mathbf{Y} \in \Gamma(\mathbf{V})$ is different from the socially optimal production, $\bar{\mathbf{Y}}$. The distortion can exist even though factors are fully and efficiently employed (e.g. distortions from variable returns to scale, imperfect markets and externalities). Second, factor-level constraints that cause shrinkage of the production possibility set owing to unemployment or inefficient employment of factors. Production constraints are denoted by β , so $\Gamma(\mathbf{V}; \beta)$ and production is chosen such that $\mathbf{Y} \in \Gamma(\mathbf{V}; \beta)$.

Denoting quantities of traded goods and factor services by the vector $\mathbf{M} = \mathbf{X} - \mathbf{Y}$ where import components of $\mathbf{M}$ are positive and export components of $\mathbf{M}$ are negative. Balanced trade balance implies that $\mathbf{P}^W \cdot \mathbf{Z} - \mathbf{A} = 0$, where $\mathbf{A}$ represents borrowing from foreigners to meet the deficit or lending to foreigners if there is a surplus. The measure of welfare change with distortion can be given as:

$$\Delta W = (\mathbf{P}^1 - \mathbf{P}^{W1}) \cdot (\mathbf{Z}^1 - \mathbf{Z}^0) + (\mathbf{A}^1 - \mathbf{A}^0) + (\mathbf{P}^{W1} - \mathbf{P}^{W0}) \cdot \mathbf{Z}^0 + \eta_p \quad (2.66)$$

Where the production effect is:

$$\eta_p = \mathbf{P}^1 \cdot (\mathbf{Y}^1 - \mathbf{Y}^0) \quad (2.67)$$

The measure of welfare change can be disaggregated into five terms. The revenue effect ($\mathbf{R}$) measures the increase in tariff revenue when the levels of imports increase from $\mathbf{Z}^0$ to $\mathbf{Z}^1$. The revenue effect is the difference between the

domestic price and world price at S^1 times the difference between imports at time $\{0, 1\}$. $\mathbf{R} = (\mathbf{P}^1 - \mathbf{P}^{W1}) \cdot (\mathbf{Z}^1 - \mathbf{Z}^0)$. If $\mathbf{T}^1 = 0$, then $\mathbf{P}^1 = \mathbf{P}^{W1}$. If there is free trade or if the level of imports does not change, the revenue effect is zero. The transfer effect ($\mathbf{F}$), is the change in transfers to or from foreigners through lending or borrowing, $(\mathbf{A}^1 - \mathbf{A}^0)$. The terms of trade effect is the change in world price $\mathbf{P}^W$ before and after, which is zero for a small country. The production effect, η_p , is the difference between the value of the final production bundle and that of the initial one. The aggregate change in welfare is the sum of these effects.

Equation 2.66 evaluates the change in welfare of the economy when moving from one situation to another. For a small country the change in welfare is:

$$\Delta W = (\mathbf{P}^1 - \mathbf{P}^{W1}) \cdot (\mathbf{Z}^1 - \mathbf{Z}^0) + (\mathbf{A}^1 - \mathbf{A}^0) + \eta_p \quad (2.68)$$

The production effect is always non-negative if, in the final situation, the production is optimal and the initial output feasible, $\mathbf{Y}^0 \in \Gamma(\mathbf{V}^1)$. Then $\mathbf{P} \cdot \mathbf{Y}^1 \geq \mathbf{P} \cdot \mathbf{Y}^0$ (this is the case if production is perfectly competitive). In general, because of the presence of production distortions, the production effects can be positive, negative or zero. The sign and magnitude depend upon substitution between goods and factors in production, good and factor prices and the constraints in the economy. If substitution in production is ruled out, the production effect is zero.

2.5.6 Conclusions

Because industry and export market diversification is hard to achieve in small economies, greater social homogeneity may diminish the associated disadvantages through flexibility and adaptability. Small countries are associated with a number of special characteristics, including a relatively large tradeables sector, dependence upon a narrow range of exports and a relatively large public sector. These characteristics arise from the small size of the domestic market, lack of natural resources and indivisibilities in public administration. Although structural openness is a consequence of their small size, functional openness is the outcome of a conscious endogenous policy choice (Demas 1965; Armstrong and Read 1998). As trade is liberalised, regional and cultural minorities can 'afford' to split because political borders do not define the size of a market; smaller countries can therefore enjoy the benefits of cultural homogeneity without suffering the cost associated with small markets (Alesina and Spolaore 2003).

The principle disadvantages associated with small size result from the existence of diseconomies of scale in various activities which inhibit the widespread attainment of international levels of competitiveness by small nations. At the same time, small states must trade-off the increased risk of domestic specialisation against the costs of diversification. International trade provides the means for small economies to overcome the inherent diseconomies of small size by extending their market (Armstrong and Read 1998).

2.6 Summary and Conclusions

This chapter reviews the theory of economic integration, the implications of non-zero trade costs and also the implications of small size for economies to provide the theoretical tools for the applied work undertaken in Chaps. 3–6.

The initial discussion of economic integration outlines the principal types of preferential trade agreements (PTAs). This is followed by a definition of economic integration and a discussion of the traditional Vinerian welfare analysis of the effects of economic integration, and a general equilibrium analysis of economic integration using Meade's incorporation of trade costs. This welfare analysis is extended by brief discussions of the theory of the second best and the transfer principle. This is followed by a discussion of trade flows and welfare effects of economic integration.

Section 2.4 provides a deeper investigation into the implications of trade costs for the welfare effects of economic integration. This discussion incorporates a brief discussion of both transportation costs and also trade barriers (tariff and non-tariff). This facilitates a discussion of the welfare implication of trade costs.

The final section presents a brief overview of the relevant literature on the economic implications of small size and the consequent effects of trade costs and economic integration. In so doing, the discussion attempts to synthesise the literature on the economic characteristics of small countries and trade costs in the light of the critical importance to both of openness to trade. The analysis then considers the implications of small size and the welfare effects of economic integration, taking issue with the commonly-held view that these effects are always positive.

Asymmetric Economic Integration
Size Characteristics of Economies, Trade Costs and
Welfare

Snorrason, S.Th.

2012, XX, 208 p., Hardcover

ISBN: 978-3-7908-2860-3

A product of Physica-Verlag Heidelberg