ANNALES D'ÉCONOMIE ET DE STATISTIQUE. – N° 59 – 2000

Externalities and Free Trade Agreements

Fritz GRAFE, Ana MAULEON *

ABSTRACT. – This paper studies the consequences of a private or depletable externality on free trade agreements in a general equilibrium framework. It is shown that there always exists a stable system of free trade spaces in the world economy. This stable system can result in either non-cooperation, partial cooperation, or total cooperation among countries of different types. The non-cooperation system is Pareto dominated by any other cooperating system. By enforcing a tax policy, a GATT arbitrator is able to implement Pareto superior stable systems in which countries of different type are cooperating.

Externalités et accords de libre-échange

RÉSUMÉ. – Dans un modèle d'équilibre général, nous analysons les conséquences d'une externalité privée sur les accords de libre-échange. Nous démontrons qu'il existe toujours un système stable d'espaces de libre-échange dans l'économie mondiale, et que tout système stable résulte soit en la non-coopération, soit en la coopération partielle, ou en la coopération totale entre pays de types différents. Il est à noter que le système stable de non-coopération est Pareto-dominé par les autres systèmes. Néanmoins, en introduisant une taxe, un médiateur du GATT est à même d'implémenter les systèmes stables Pareto-supérieurs où des pays de types différents coopèrent.

* F. GRAFE: Institute of Public Economics, Basque Country Univeristy ;
A. MAULEON: Department of Economic Analysis, Basque Country University.
We thank E. INARRA, O. VOLIJ, D. PEREZ CASTRILLO and two anonymous referees for their comments and suggestions. A special thank goes to V. VANNETELBOSCH for helpful discussions and comments. This research has been supported by the projects DGICYT PB94-1372, PI98/48 and UPVI01.I01-G17/99.

1 Introduction

A free trade agreement between developed and developing countries may involve not only movements of production factors or consumption goods, but also may involve transfers of negative externalities. Low and Yeats [1992] found that between 1965 and 1988 the 4.5 percent of dirty industries moved from industrial to non-industrial countries, mainly from North to South America, thanks to capital movements between these two areas under the form of foreign direct investment.[1]

One class of negative externalities are private or depletable externalities. Baumol and Oates [1988] defined a private or depletable externality (acid rain, trash dumped on private properties, etc) as the one in which one victim's consumption of the externality reduces that of others. The kind of externality we have in mind is, for example, the depletable waste (like trash) emitted by the factories located in each country that is supported by the inhabitants of the emitting country without affecting other countries.

Up to now, the question of whether the traditional gains from trade are sufficient to offset the loss of environmental quality has not been addressed in the literature on international economic integration. Using a cooperative game-theoretic approach we study, in terms of welfare considerations, the impact of such depletable externalities on the formation and stability of free trade agreements in a general equilibrium framework.

We consider a world economy formed by two types of countries having the same welfare function, the same endowment of labor but having different endowment of capital.[2] The technology, available to any coalition of countries, transforms capital and labor into a single consumption good. It is assumed that the production in a country generates a negative externality on the welfare of that country, but not on the welfare of other countries. The greater the capital endowment of a country, the greater its production and its consumption level, but also the greater the negative externality it has to suffer.

The classical consequence of free mobility of factors and good is the equalization of factor and good prices and the increase in the world production (see *e.g.* Krugman and Obstfeld [1991]). Apart from this classical effect, free mobility of factors allows countries to transfer part of their production as well as part of their negative externality to other countries. The welfare function of any country depends positively on its consumption level and negatively on its production level. Therefore, countries with relatively high capital endowment reduce their negative external effect by transferring part of their resources to countries with relatively low capital endowment. At the same time, they benefit from a greater consumption level due to the increase in the world

1. Moreover, Lucas *et al.* [1992] found that the fall in the pollution intensity of national product in many developed countries is due to a change in the composition of output and not a movement toward cleaner production methods.
2. Dagan and Volij [1997] have used a similar general equilibrium model to study the effects of a redistributive policies on the formation of nations in an economy where people have preferences for income redistribution.

production. On the contrary, free mobility of factors has an ambiguous effect on the welfare of countries with relatively low capital endowment. For example, they are worse off if the increase of their negative external effect dominates the increase of the positive effect due to higher consumption levels. Therefore, countries with relatively low capital endowment could prefer not cooperating with countries with relatively high capital endowment in the presence of depletable externalities, while they would be interested in cooperating in the absence of depletable external effects.

Hence, in the absence of depletable externalities, the cooperation among all the countries or total cooperation is Pareto superior to partial or non-cooperation in world welfare terms. That is, in the absence of externalities, our model confirm the classical assertion of international trade theory.[3] Is total cooperation still Pareto efficient once depletable external effects are present?

To answer this question, we derive the equilibrium welfare that each country obtains when it participates in a free trade space with countries of different type, and we study the stability of these cooperation agreements. The definition of stability we use is the following one. A partition of the world economy into free trade spaces is *stable* if no group of countries would be better off by forming a new free trade space.

The choice of the core to describe the stable coalition structure is justified, as in KOWALCZYK and SJOSTROM [1994], thanks to the absence of strategic externalities between coalitions. KOWALCZYK and SJOSTROM [1994] used a many-country model of world trade to investigate the existence and character of a global financial mechanism that, in conjunction with the elimination of all distortionary trade practices, brings global free trade into the core.[4]

In our model, the absence of strategic externalities between coalitions is due to the definition of a free trade space. As in LEVY [1997], a free trade space consists of two or more countries setting tariffs between them to zero and a prohibitive common tariff to the rest of the world so as to induce autarchy for each space.[5] Hence, the welfare of each country participating in a free trade space does not depend on the coalition structure formed. The lack of external tariffs different than the prohibitive one, allows us to focus on the choice of free trade spaces and to avoid the difficult question of how external tariffs might be determined simultaneously.[6] Another branch of the literature on international trade has focused either on welfare effects of trading blocs: KRUGMAN [1991] showed that, in a model of monopolistic competition, welfare is U-shaped in the number of trading blocs; or on stability questions:

---

3. Under competitive conditions, free trade between different types of countries is Pareto superior to autarky under some redistributive measures (see, DIXIT and NORMAN [1980]).
4. Other related papers have considered the notion of the core. RIEZMAN [1985] in a three-country, three-good pure exchange model without inter-country tranfers found distributions of initial endowments for which the multilateral free trade agreement is not in the core, while a customs union is in the core. MACHO-STADLER *et al.* [1998] have analysed multilateral tariff negotiations among three countries as a game in coalition form.
5. LEVY [1997], assuming that tariffs are either zero or prohibitive, has shown that bilateral free trade agreements can undermine support for multilateral free trade in a differentiated-product model with variety gains but not in a Heckscher-Ohlin setting.
6. In the usual definition of a customs union (RIEZMAN [1985]), the common external tariff is set jointly to maximize the aggregate welfare of members. Therefore, the volume of trade of a customs union depends on the outer tariff set with respect to outside countries.

KENNAN and RIEZMAN [1990] constructed a model of customs unions in which countries set optimal tariffs and they showed that if the customs union is big enough it can improve its members' welfare over free trade; BOND and SYROPOULOS [1996] proved the instability of a symmetric customs union structure in a pure exchange model of trade among welfare-maximizing countries with CES preferences. But all these contributions take the structure of trading blocs as exogenously given.

Our main results are the following ones. Firstly, there always exists a stable system of free trade spaces in the world economy. We have identified five different stable systems of free trade spaces. These stable systems result in either non-cooperation, partial cooperation, or total cooperation among countries of different type. Secondly, from the point of view of the world welfare, the non-cooperation situation is Pareto dominated by any other cooperative situation between countries of different type.

Several previous studies have focused on the impact of trade liberalization, in presence of external effects, on welfare levels.[7] In a more closely related paper, COPELAND and TAYLOR [1994] have developed a two-country general equilibrium model with a continuum of goods differing in their pollution intensity of production. Under the assumptions that pollution is confined to the emitting country and that governments regulate pollution optimally, they showed that trade liberalization is always welfare improving even when it raises pollution levels. In their model, all pollution-intensive industries locate in the South (the capital poor country) while the relatively clean industries locate in the North (the capital rich country) due to the fact that North chooses a higher pollution tax. Hence, trade always lowers the pollution level in the North, increases the pollution level in the South, and increases worldwide pollution provided that factor prices are not equalized across countries. Finally, COPELAND and TAYLOR [1995] have investigated the opposed case of transboundary pollution emissions but in a world formed now by countries of two different types differing in their human-capital levels. Since pollution crosses borders, uncoordinated regulation of pollution at the national level does not eliminate all market failures, and consequently free trade does not raise welfare. In a factor prices equalized equilibrium world pollution is unaffected, while in a factor prices not equalized equilibrium world pollution raises.

In our model, although pollution stayed within the country of origin, there is market failure since national governments do not regulate the private externality. As a result, the Pareto dominated system of free trade spaces in which countries of different type do not cooperate could be stable. Which kind of environmental policy could be used by the national governments to regulate this private externality?[8] Since the OECD (Organization for Economic

7. MARKUSEN [1975] determined necessary conditions for an optimal tax structure in order to maximize social welfare in a model of two-commodity and two-trading countries related by a bilateral production externality. In a two-good Ricardian model with pollution, PETHIG [1976] showed that the country which allows a higher emissions' level will export the pollution-intensive good. SIEBERT *et al.* [1980] investigated the impact of a national emission tax in a two-country, two-commodity and two-factor world with capital movements.

8. As FREEMAN [1984] has shown, the distinction between public and private externalities does not have any fundamental implications for the pricing of externalities: a Pigouvian levy on the generator equal to marginal social damage and no compensation or taxes on the victims is necessary to reach Pareto optimality.

Cooperation and Development) countries adopted the polluter pays principle in 1972 as a guideline for domestic environmental policies, all the conventions, protocols and agreements signed on the basis of the Stockholm Declaration presume the polluter pays principle (see, MALER [1990]). Standards and taxes are the two environmental policies compatible with the polluter pays principle according to GATT (General Agreement on Tariffs and Trade) rules (PEARCE [1995]). Given our assumptions that all countries have the same environmental standards and that the pollution is confined to the country of origin, a tax on the production should be levied.[9] Assume that a free trade agreement involves an International Environmental Agreement (IEA) in which all participating countries should agree on the uniform tax production rate to be levied. In order to avoid the discussion about how an actual process of negotiating countries submitting offers and counteroffers would or should work, we assume that the participating countries ask the mediation of a GATT arbitrator. This GATT arbitrator, having or not information about the welfare functions of the participating countries, can cut short any time-consuming bargaining process by presenting an uniform tax production rate that will be accepted by all the participating countries. Moreover, enforceability of the agreement reached is guaranteed given that the collecting of taxes and the movements of capital take place simultaneously.

Finally, in relation with the role of the GATT arbitrator, we provide an optimal tax policy according to the polluter pays principle which maximizes the world welfare level when all countries in the world economy are cooperating into a single free trade space.

The paper is organized as follows. In Section 2, the general equilibrium model is presented. In Section 3, after introducing the definition of stability, we show the existence of a stable system of free trade space for the world economy and we characterize all the possible stable systems. Section 4 is devoted to the intervention of a GATT arbitrator. Section 5 concludes.

2 The Model

In this section, we develop a model that allows us to analyze the stability of free trade agreements between two types of countries when there exists a private negative externality in the production.

Let E be the set of all countries in the world. Each country is characterized by a technology, F, a welfare function, U, and an endowment of two production inputs, capital, K, and labor, L. All countries have access to the same technology $F : \mathbb{R}_+^2 \to \mathbb{R}_+$ that transforms capital and labor into a consumption good x.

9. In MARKUSEN [1975], when the externality is reduced to a simple domestic distortion, a production tax represents a first best solution.

ASSUMPTION 1: The production function F is strictly quasi-concave, exhibits constant returns to scale and satisfies

$$\lim_{K \to 0} \partial F / \partial K = \infty, \ \lim_{K \to \infty} \partial F / \partial K = 0,$$

and for all $K > 0$, $\partial F / \partial K > 0$ and $\partial^2 F / \partial K^2 < 0$.

Let $U(x, F(K, L))$ be the welfare function of each country. In order to study the impact of production negative externalities on the formation and stability of free trade agreements, we choose a welfare function which depends positively on the amount of the good, x, and negatively on the production level.[10]

ASSUMPTION 2: $U(x, F(K, L)) = x - v(F(K, L))$, where v is a strictly increasing and convex function.

Since the objective of the paper is to analyze the effects of a negative externality on free trade agreements between different types of countries, we separate the world economy E in two groups of countries[11] differentiated only by their capital endowment: $E = E_1 \cup E_2$, with $E_1 \cap E_2 = \emptyset$, where E_i is the set of countries with capital endowment K_i, $i = 1, 2$. Let $K_2 > K_1$. Also, we normalized the labor endowment of each country: $L_i = 1$, $i = 1, 2$.

Before formulating the equilibrium notion, we define a free trade space (FTS) and we introduce some additional notation.

DEFINITION 1: A free trade space is a disjoint pair $C = (C_1, C_2)$ where $C_1 \subseteq E_1$ and $C_2 \subseteq E_2$, such that between the members of C_1 and C_2 there exist a free trade of inputs and output.

Definition 1 allows that a free trade space C can also be formed by countries of a unique type; *e.g.* $C = (C_1, \emptyset)$ or $C = (\emptyset, C_2)$. Given any collection of free trade spaces $\langle C^1, C^2, \ldots \rangle$ where $C^j = \left(C_1^j, C_2^j\right)$, let

$$\bigcap_j C^j \equiv \left(C_1^1 \cap C_1^2 \cap \ldots, C_2^1 \cap C_2^2 \cap \ldots\right)$$

and $\bigcup_j C^j \equiv \left(C_1^1 \cup C_1^2 \cup \ldots, C_2^1 \cup C_2^2 \cup \ldots\right)$.

For any free trade space $C \subseteq E$, let $c_i = \lambda(C_i)$ denotes the measure of the set C_i (for $i = 1, 2$), where λ is a well-defined measure. Then, $K(C) = c_1 K_1 + c_2 K_2$ is the capital used in the free trade space C to produce x, and $L(C) = c_1 + c_2$ is the labor in C. Since the technology exhi-

10. The assumption that all countries have the same welfare function implies that all of them have the same environmental standard.

11. We restrict our analysis to two types of countries since this simple case captures most important effects and results that would be obtained with three or more different types of countries.

bits constant returns to scale, the total production in C is equal to

$$F(K(C), L(C)) = f(k(C)) \cdot (c_1 + c_2),$$

where $f(k(C))$ is the per country production function in C,

and $$k(C) = [c_1 K_1 + c_2 K_2] / [c_1 + c_2]$$

is the capital-labor ratio in C. Let $\overline{k}$ denotes the capital-labor ratio of the world economy E. That is, $\overline{k} = k(E) = [e_1 K_1 + e_2 K_2] / [e_1 + e_2]$, where $e_i = \lambda(E_i)$ is the measure of the set E_i (for $i = 1,2$).

Since in a free trade space C there exists free mobility of inputs and output, the prices must be the same in all the countries belonging to C. We denote by r and w the common market prices of capital and labor, respectively, in C, and p_x the price of the good x. Given the prices p_x, r and w, let $z_x(p_x, r, w; C)$, $z_K(p_x, r, w; C)$ and $z_L(p_x, r, w; C)$ be the excess demand functions in the FTS C. Then, the *equilibrium prices* in the free trade space C are:

$$\text{(1)} \qquad \begin{cases} p_x^* = 1 \\ r^* = f'(k(C)) \\ w^* = f(k(C)) - f'(k(C))\, k(C) \end{cases}$$

because for $r^* = f'(k(C))$ and $w^* = f(k(C)) - f'(k(C))\, k(C)$ the factors market clearing conditions are satisfied: $z_K\left(p_x^*, r^*, w^*; C\right) = 0$ and $z_L\left(p_x^*, r^*, w^*; C\right) = 0$. Moreover, the consumption of the good x in each country of type $i \in C_i$ equals the retributions of capital and labor in that country: $x_i = rK_i + w$ (with $p_x = 1$), for $i = 1,2$. Then, the total consumption of good x in C is equal to the total production in C:

$$c_1 x_1 + c_2 x_2 = f(k(C)) \cdot (c_1 + c_2); \text{ and, } z_x\left(p_x^*, r^*, w^*; C\right) = 0.$$

Substituting the equilibrium prices in the welfare function, we obtain the indirect welfare function of a country of type $i \in C_i$ (for $i = 1,2$):

$$\text{(2)} \qquad \begin{aligned} W_i(k(C), K_i) &= x_i - v(f(k(C))) \\ &= f(k(C)) - f'(k(C))\,(k(C) - K_i) - v(f(k(C))) \end{aligned},$$

which depends only on his own endowment of capital K_i and on the capital-labor ratio $k(C)$.[12] This function gives us the equilibrium welfare enjoyed by a country of type $i \in C_i$ when it participates in the free trade space C. However, if a country of type i does not participate in a free trade space with countries of different type, then the welfare level of country of type i is equal to $W_i(K_i) = f(K_i) - v(f(K_i))$.

Before analyzing the stability of free trade spaces, we state some properties of the indirect welfare function W_i.

12. The indirect welfare function does not depend on the coalition structure formed since, in the definition of a free trade space, we are assuming that countries participating in a free trade space set a prohibitive common tariff to the rest of the world so as to induce autarky for each space.

(i) The function W_i is increasing in the capital endowment K_i:

$$\partial W_i / \partial K_i = f'(k(C)) > 0.$$

(ii) Since $\partial W_i / \partial k(C) =$

$$-f''(k(C))(k(C) - K_i) - v'(f(k(C)))\, f'(k(C)),$$

a change in the capital-labor ratio of the FTS has two effects.[13] On the one hand, an increase in the capital-labor ratio reduces the equilibrium price of capital, increasing the income of those countries with relatively low capital endowment and decreasing the income of those countries with relatively high capital endowment. On the other hand, the increase in the capital-labor ratio raises the production per country, $f(k(C))$, affecting negatively the welfare of each country in C.

These two properties (i) and (ii) imply that an increase in $k(C)$ strictly reduces the welfare of countries with relatively high capital endowment but has an ambiguous effect on the welfare of countries with relatively low capital endowment. In other words, although countries in E_2 would like to form any free trade space with countries in E_1 (because when the capital-labor ratio decreases, it affects positively their income and it reduces the negative externality they suffer), countries in E_1 may prefer not to participate in that spaces (because an increase in the capital-labor ratio affects positively their income but it also increases the negative externality they suffer).

Therefore, the following interesting question arises. Does there exist a partition of the world economy into *stable* free trade spaces? In the next section, we will answer this question.

3 Stable System of Free Trade Spaces

3.1 Existence

Before proving that there always exists a stable partition of the world economy into FTS, we have to define the stability notion. A system of free trade spaces (FTS) is *stable* if no group of countries can benefit by forming a new free trade space. Let us define it formally.

13. In the absence of the external negative effect, the expression

$$\partial W_i / \partial k(C) = -f''(k(C))(k(C) - K_i)$$

would be negative for countries in E_2 and positive for countries in E_1. In other words, it would be advantageous for both types of countries to form a free trade space with countries of different types.

DEFINITION 2: A system of FTS is a finite collection of disjoint free trade spaces $\mathcal{S} = \langle C^1, C^2, \ldots, C^n \rangle$ such that $\bigcup_{j=1}^{n} C^j = E$.

DEFINITION 3: A system of FTS $\mathcal{S} = \langle C^1, C^2, \ldots, C^n \rangle$ is *stable* if there does not exist an equilibrium for a free trade space T such that for every country of type $i \in T_i$ (for $i = 1, 2$): $W_i\,(k\,(T)\,, K_i) > W_i\left(k\left(C^j\right), K_i\right)$ with $i \in C_i^j,\ C^j \in \mathcal{S}$.

Definition 3 is our stability notion. It simply means that a partition of the world economy into free trade spaces is stable if it is not possible to form a new free trade space, in which all the participating countries are better off at equilibrium. Note that without externality, the indirect welfare function is increasing in $k\,(C)$ for countries with relatively low capital endowment, and decreasing in $k\,(C)$ for countries with relatively high capital endowment. In this case, the free trade space formed by all countries in the world economy is stable.

Notice that the welfare of a country of type $i \in C_i$, $W_i\,(k\,(C)\,, K_i)$, does not depend on the measure of the FTS C. Then, we can join (or split) some FTS with the same capital-labor ratio in such a way that, in this new FTS, all countries obtain the same welfare than before. In what follows, we will consider only systems of FTS in which all FTS have different capital-labor ratios (and hence, different welfare level).

In order to prove the existence of stable systems of FTS, some more notation is needed. Given any FTS C, let

$$\overline{K}_i = \{k \in [K_1, K_2] : W_i\,(k\,(C)\,, K_i) \geqslant W_i\,(K_i)\},$$

for $i = 1, 2$. $\overline{K}_i$ is the set of capital-labor ratios such that a country of type $i \in C_i$ obtains a welfare greater or equal than the guaranteed welfare (individually rational condition). By guaranteed, we mean the welfare a country can obtain under autarchy or under cooperation with countries of its own type. Let $\overline{K} = \overline{K}_1 \cap \overline{K}_2$ be the set of capital-labor ratios individually rational for both types of countries. Let $K_i^{\max} = \{k \in [K_1, K_2] : k = \arg\max W_i\,(k\,(C)\,, K_i)\}$, for $i = 1, 2$. It is the set of capital-labor ratios such that a country of type $i \in C_i$ obtains its maximum welfare.

Since $W_2\,(k\,(C)\,, K_2)$ increases when $k\,(C)$ decreases, we have

$$\overline{K}_2 = [K_1, K_2] \supset \overline{K}_1 = \overline{K} \text{ and } K_2^{\max} = \{K_1\}.$$

That is, the set of capital-labor ratios individually rational for both types of countries is $\overline{K}_1$. Also, each country of type 2 obtains its maximum welfare participating in a FTS with countries of type 1 altogether.

THEOREM 1: There always exists a stable system of FTS for the world economy E.

FIGURE 1

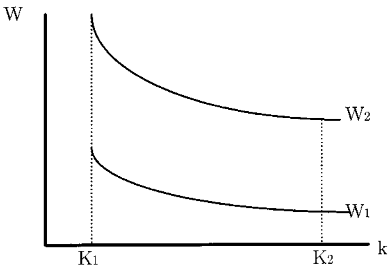


FIGURE 2

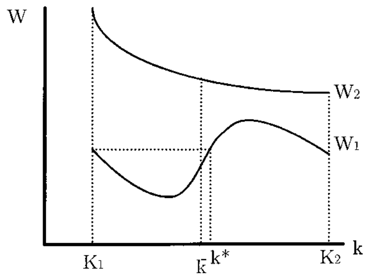


FIGURE 3

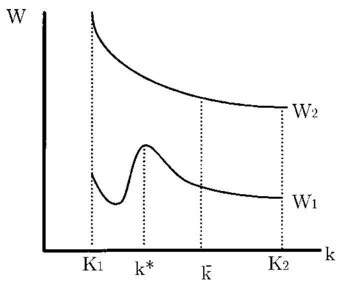


FIGURE 4

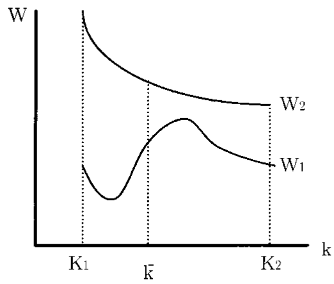
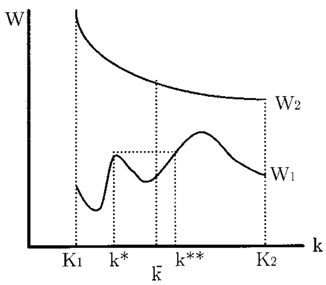


FIGURE 5



The proof of Theorem 1, as well as the other proofs not in the main text, can be found in the Appendix. Next, we give the reader the intuition behind our proof. We do it graphically and for the case where $K_1^{\max} = \{k_1^{\max}\}$ is a singleton. In this case, the uniqueness of the equilibrium structure is guaranteed. We can distinguish two main cases.

1. There is no capital-labor ratio individually rational for both types of countries in the interval $(K_1, K_2]$: $\overline{K} = \{K_1\}$. In this case, countries of type 1 are better off not cooperating with countries of type 2. This case is depicted in Figure 1. The unique stable system of FTS, $\mathcal{S}_0$, is such that all countries of the same type remain together: $\mathcal{S}_0 = \langle C^1, C^2 \rangle$ with $C^1 = (E_1, \emptyset)$ and $C^2 = (\emptyset, E_2)$.

2. There exist capital-labor ratios individually rational for both types of countries in the interval (K_1, K_2): $\overline{K} \neq \{K_1\}$. Three subcases have to be considered.

2.a. The capital-labor ratio of the world economy $\overline{k}$ is smaller than any capital-labor ratio k $(k \neq K_1)$ individually rational for both types of countries: $\overline{k} < k$, $\forall k \in \overline{K}$, $k \neq K_1$. It follows immediately that the candidates FTS, where cooperation occurs, are such that the measure of countries of type 2 is greater than the measure of countries of type 1. In the unique stable system of FTS, $\mathcal{S}_1$, all countries of type 2 form a FTS, C^2, with the optimal measure of countries of type 1 in order to maximize their welfare. This optimal measure is such that countries of type 1 are indifferent between belonging to the FTS, C^2, and remaining with only countries of their own type. Thus, we have $\mathcal{S}_1 = \langle C^1, C^2 \rangle$ where $c^1 = \lambda\left(C^1\right) < e_1$ with $k\left(C^1\right) = K_1$ and

$$c^2 = \lambda\left(C^2\right) = \left(e_1 - c^1\right) + e_2$$

with $k\left(C^2\right) = k^* = \arg\max_{k \in \overline{K}} W_2\left(k\left(C\right), K_2\right)$ (see, Figure 2). The system $\mathcal{S}_1$ is stable because countries of type 1 would like to form FTS with capital-labor ratios $k > k^*$, but no country of type 2 would like to participate in that FTS. Moreover, countries of type 2 would like to form FTS with capital-labor ratios $k < k^*$, but no country of type 1 would like to participate in that FTS.

2.b. The capital-labor ratio of the world economy $\overline{k}$ is greater than any capital-labor ratio k individually rational for both types of countries: $\overline{k} > k$, $\forall k \in \overline{K}$. It follows immediately that the candidates FTS, where cooperation occurs, are such that the measure of countries of type 1 is greater than the measure of countries of type 2. In the unique stable system of FTS, $\mathcal{S}_2$, all countries of type 1 form a FTS, C^2, with the optimal measure of countries of type 2 in order to maximize their welfare. Thus, we have $\mathcal{S}_2 = \langle C^1, C^2 \rangle$ where $c^1 = \lambda\left(C^1\right) < e_2$ with $k\left(C^1\right) = K_2$ and $c^2 = \lambda\left(C^2\right) = \left(e_2 - c^1\right) + e_1$ with $k(C^2) = k^*$ $= \arg\max_{k \in \overline{K}} W_1\left(k\left(C\right), K_1\right) = k_1^{\max}$ (see, Figure 3). This system $\mathcal{S}_2$ is stable because countries of type 2 would like to form FTS with capital-labor ratios $k < k^*$, but no country of type 1 would like to participate in that FTS. Moreover, countries of type 1 get their maximum welfare level.

2.c. $\min\{k : k \in \overline{K}, k \neq K_1\} < \overline{k} < \max\{k : k \in \overline{K}\}$. Both types of free trade spaces described in (2.a) and (2.b) are possible. We distinguish three different classes of stable systems of FTS. The first class is identical to (2.b). In the second class, all countries of the world economy form a stable free trade space: $\mathcal{S}_3 = \langle E \rangle$. This system $\mathcal{S}_3$ is stable because any potential FTS that countries of type 1 would like to form with $k > \overline{k}$, needs some countries of type 2, but no country of type 2 would agree to participate in such FTS (see, Figure 4). And the same argument applies for any potential FTS that countries of type 2 would like to form with $k < \overline{k}$. Finally, in the third class, a stable system $\mathcal{S}_4$ is formed by two free trade spaces with capital-labor ratios of k^* and k^{**} respectively, in which countries of type 1 have the same equilibrium welfare level. This equilibrium welfare level is the one associated with the best individually rational capital-labor ratio for the countries on the left of $\overline{k}$ (see, Figure 5). Formally, $\mathcal{S}_4 = \left\langle C^1, C^2 \right\rangle$ where $c^1 = \lambda\left(C^1\right) = \lambda(C_1^1) + \lambda(C_2^1)$ with $k\left(C^1\right) = k^* < \overline{k}$, and $c^2 = \lambda\left(C^2\right) = (e_1 - \lambda(C_1^1)) + (e_2 - \lambda(C_2^1))$ with $k\left(C^2\right) = k^{**} > \overline{k}$. This system is stable because countries of type 1 would like to form FTS with $k > k^{**}$, but no country of type 2 would like to participate in that FTS. The same argument applies for any potential FTS that countries of type 2 would like to form.

Therefore, we have identified two types of stable systems of FTS: systems of FTS in which there is no cooperation between different types of countries ($\mathcal{S}_0$), and systems of FTS in which there is partial or total cooperation between different types of countries ($\mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3, \mathcal{S}_4$).

3.2 **World Welfare Levels**

Up to now, we have analyzed the stability of systems of FTS based on rational decisions taken, separately, by each country. In order to study the consequences of these stable systems of FTS in world welfare terms, we consider an utilitarist world welfare function, $V : \mathbb{S} \to \mathbb{R}_+$ such that

$$\mathcal{S} \longmapsto V(\mathcal{S}) = \sum_{C^j \in \mathcal{S}} \lambda\left(C_1^j\right) W_1\left(k\left(C^j\right), K_1\right) + \sum_{C^j \in \mathcal{S}} \lambda\left(C_2^j\right) W_2\left(k\left(C^j\right), K_2\right),$$

where $\mathbb{S}$ is the set of all systems of FTS. The next proposition compares the world welfare levels associated to our five stable systems of FTS.

PROPOSITION 1: Given the utilitarist world welfare function V, the world welfare of any stable system of FTS in which countries of different type cooperate is greater than the world welfare of the stable system of FTS in which countries of different type do not cooperate.

In our model, free trade of good and factors increases world pollution as a result of the increase in world production. However, if the gains due to higher consumption levels in the capital poor countries are greater than the losses due to the increase in pollution, any partial or total free trade agreement between different types of countries implies a greater welfare level for both types of countries, compared to the situation in which countries of different type do not cooperate.

This results holds due to the assumption of the utilitarist world welfare function. If instead one assume an egalitarian world welfare function according to Rawls criterium, only the stable systems represented in Figure 4 (global free trade) and Figure 5 (in which there is two free trade spaces formed by countries of both types) strictly Pareto dominate the non-cooperation situation.

4 Intervention of a GATT Arbitrator

4.1 Why a GATT Arbitrator May Be Useful?

In the previous section, we have shown that the world welfare level of any stable system of FTS in which there exists some free trade space formed by both types of countries, is greater than the world welfare level when countries of different type do not cooperate. However, the cooperation between countries of different type is not always possible, as the following example illustrates.

Example 1

Consider a world economy E formed by the same measure of countries of each type. The technology is given by the following production function $x = F(K_i, L_i) = K_i^{0.25} L_i^{0.75}$, $i = 1,2$. The initial endowments of capital and labor are $L_1 = L_2 = 1$, $K_1 = 10$ and $K_2 = 50$ and the welfare function is given by $U_i(x_i, F(K_i, L_i)) = x_i - \frac{1}{10}[F(K_i, L_i)]^2$, $i = 1,2$.

In this example, $W_1(K_1) = f(K_1) - \frac{1}{10}[f(K_1)]^2 = 1.46$,

$$W_2(K_2) = f(K_2) - \frac{1}{10}[f(K_2)]^2 = 1.95,\ \bar{k} = 30, \text{ and}$$

$$W_i(k(C), K_i) = f(k(C)) - f'(k(C))(k(C) - K_i) - \frac{1}{10}[f(k(C))]^2,$$

for $i = 1,2$, is the welfare corresponding to the equilibrium allocation when countries of both types form a free trade space C. It can be computed that for any $k(C)$ we have $W_1(k(C), K_1) < W_1(K_1)$. Then, $\overline{K} = \{K_1\}$. It implies that the unique stable system of FTS is $\mathcal{S}_0 = \langle C^1, C^2 \rangle$ with $k(C^1) = K_1$ and $k(C^2) = K_2$. ■

In our model, the Pareto dominated system of FTS in which countries of different type do not cooperate could be stable due to the fact that national governments do not regulate the private externality. Given that pollution is confined to the emitting country and that all countries have the same environmental standards, a tax on the production should be levied to regulate the externality. Assume that, instead to sign an IEA, all countries in the world economy ask the intervention of a GATT arbitrator to establish the uniform tax production rate. Two kind of interventions are considered here. In the first one, the GATT arbitrator does not know the welfare functions of the countries and chooses an exogenous tax production rate in such a way that at least a group of countries of different type cooperates (Sections 4.2 and 4.3). In the second one, the GATT arbitrator knows the welfare function of any country and chooses, according to the *polluter pays principle*, an optimal tax production rate equal to the marginal disutility caused by the externality in each country participating in a given free trade space (Section 4.4).

4.2 The Model with a GATT Arbitrator

We show how the intervention of a GATT arbitrator can reach the cooperation between both types of countries by establishing, for each possible FTS, a uniform tax production rate $t > 0$ exogenously. The new price of x is $p_x = (1 - t)$. Moreover, we assume that the amount of taxes collected in each country is redistributed among the agents living in that country. Therefore, in any country belonging to a FTS, C, the consumption of x is equal to the retributions of capital and labor plus the amount of taxes collected in that country. That is, $x_i = rK_i + w + tf\,(k\,(C))$, for $i = 1{,}2$.

Given the tax production rate $t > 0$, the equilibrium prices in the FTS C are now:

$$(3) \qquad \begin{cases} p_x^* = (1 - t) \\ w^* = (1 - t)\left[f\,(k\,(C)) - f'\,(k\,(C))\,k\,(C)\right] \\ r^* = (1 - t)\; f'\,(k\,(C)) \end{cases}$$

Substituting the new equilibrium prices in the welfare function, we obtain the new indirect welfare function for a country of type $i \in C_i$ (for $i = 1{,}2$):

$$(4) \qquad \begin{aligned} W_i\,(k\,(C)\,,K_i,t) &= x_i - v\,(\,f\,(k\,(C))) \\ &= f\,(k\,(C)) - (1 - t)\; f'\,(k\,(C))\,(k\,(C) - K_i) \\ &\qquad - v\,(\,f\,(k\,(C))), \end{aligned}$$

which depends on his own endowment of capital, on the capital-labor ratio in the FTS C, and on the tax production rate chosen by the GATT arbitrator.

Before analyzing the stability of systems of FTS under a GATT arbitrator, we give some properties of the indirect welfare function $W_i\,(k\,(C)\,,K_i,t)$.

(i) The derivative $\partial W_i/\partial t = f'\,(k\,(C))\,(k\,(C) - K_i)$ is positive for countries in E_1 because they collect $t\;f\,(k\,(C))$ and the net income of the owners of capital and labor only decreases in

$$t \cdot [\,f\,(k\,(C)) - f'\,(k\,(C))\,(k\,(C) - K_1)],$$

but is negative for countries in E_2. In fact, by establishing a tax production rate the GATT arbitrator is implementing income transfers from the rich to the poor countries in order to guarantee the participation of the capital poor countries in any FTS.

(ii) The function W_i is increasing in the capital endowment K_i. Indeed, we have that $\partial W_i / \partial K_i = (1-t) f'(k(C)) > 0$.

(iii) A change in the capital-labor ratio of the FTS has three effects. Two effects are already present in the case without GATT arbitrator (see (ii) of the properties of the indirect welfare function for the case without GATT arbitrator). But another effect appears with the intervention of a GATT arbitrator. The increase of $f(k(C))$ when $k(C)$ increases has a positive effect on the amount of taxes collected in each country participating in C. As a consequence of this effect, an increase in $k(C)$ has an ambiguous effect on the welfare level of both types of countries:

$$\partial W_i / \partial k(C) = t f'(k(C)) - (1-t) f''(k(C))(k(C) - K_i) \\ - v'(f(k(C))) f'(k(C)) \gtrless 0.$$

4.3 **Existence of Stable FTS under an Exogenous Tax Policy**

Before we give some conditions under which the GATT arbitrator can increase the world welfare level by means of the exogenous tax policy, we introduce more notations.

Given any FTS C and the tax production rate $t > 0$, the set of capital-labor ratios such that a country of type $i \in C_i$ obtains a welfare greater or equal than the guaranteed welfare becomes

$$\overline{K}_i(t) = \{k \in [K_1, K_2] : W_i(k(C), K_i, t) \geqslant W_i(K_i)\},\ i = 1,2.$$

Let $\overline{K}(t) = \overline{K}_1(t) \cap \overline{K}_2(t)$ be the set of capital-labor ratios individually rational for both types of countries when the tax production rate is t. Let $\overline{K}_i^{\max}(t) = \left\{k \in \overline{K}(t) : k = \arg\max W_i(k(C), K_1, t)\right\},\ i = 1,2$ be the set of capital-labor ratios individually rational for both types of countries and such that a country of type $i \in C_i$ obtains its maximum welfare.

We say that countries of type 1 (type 2) are *cooperation averse* if $k < \overline{k}$, $\forall k \in \overline{K}_1^{\max}(t)$ (if $k > \overline{k}, \forall k \in \overline{K}_2^{\max}(t)$), *i.e.* if all capital-labor ratios that maximize its welfare (and individually rational) are smaller (greater) than the capital-labor ratio of the world economy. Also, countries of type 1 (type 2) are *cooperation lover* if $k > \overline{k}$, $\forall k \in \overline{K}_1^{\max}(t)$ (if $k < \overline{k}, \forall k \in \overline{K}_2^{\max}(t)$), *i.e.* if all capital-labor ratios that maximize its welfare (and individually rational) are greater (smaller) than the capital-labor ratio of the world economy.

Notice that the sets $\overline{K}_i^{\max}(t)$ are determined by the welfare function, the technology and the endowments of capital. These sets can contained more than one point. However, we now assume that the sets $\overline{K}_i^{\max}(t)$ are singletons : $\overline{K}_i^{\max}(t) = \{k_i^{\max}\},\ i = 1,2.$

Let $\widehat{k}_1 = \underset{k \in \overline{K}(t), k \leqslant \overline{k}}{\arg\max}\ W_1 (k (C), K_1, t)$ be the capital-labor ratio in the interval $[K_1, \overline{k}\,]$ such that countries of type 1 obtain their maximum welfare. Let $\widehat{k}_2 = \underset{k \in \overline{K}(t), k \geqslant \overline{k}}{\arg\max}\ W_2 (k (C), K_2, t)$ be the capital-labor ratio in the interval $[\,\overline{k}, K_2]$ such that countries of type 2 obtain their maximum welfare.

In order to prove the existence of stable systems of FTS, we impose the following two assumptions.

ASSUMPTION 3: The tax production rate $t > 0$ is such that

$$\overline{K}(t) \cap (K_1, K_2) \neq \emptyset.$$

ASSUMPTION 4: If both types of countries are « cooperation lover » then $W_1(\widehat{k}_2, K_1, t) \geqslant W_1(\widehat{k}_1, K_1, t)$ and $W_2(\widehat{k}_1, K_2, t) \geqslant W_2(\widehat{k}_2, K_2, t)$.

Assumption 3 is necessary to justify the intervention of the GATT arbitrator. On the contrary, if there does not exist a tax production rate $t > 0$ such that $\overline{K}(t) \cap (K_1, K_2) \neq \emptyset$, there would not be a margin for the intervention of a GATT arbitrator. In Example 1, if $t = 0.01$ then $\overline{K}(t) = \{K_1\}$, but $\overline{K}(t) \cap (K_1, K_2) \neq \emptyset$ if $t = 0.20$. The existence of a tax production rate $t > 0$ such that $\overline{K}(t) \cap (K_1, K_2) \neq \emptyset$ depends on the welfare function, the technology and the initial endowments. However, if we assume that the welfare function of any country of type i being alone,

$$W_i (K_i) = f (K_i) - v (f (K_i)),$$

is a concave function for all K_i, the existence of such tax production rate will be guaranteed.

Assumption 4 can be interpreted in the following way. In the interval $[K_1, \overline{k}\,]$ $([\,\overline{k}, K_2])$ any FTS contains a measure of countries of type 1 (type 2) greater than the measure of countries of type 2 (type 1). Then, the cooperation in the interval $[K_1, \overline{k}]$ $([\,\overline{k}, K_2])$ is only possible if countries of type 1 (type 2) obtain the welfare level $W_1(\widehat{k}_1, K_1, t)$ $(W_2(\widehat{k}_2, K_2, t))$. Also, countries of type 1 (type 2) participate in a FTS in the interval $[\,\overline{k}, K_2]$ $([K_1, \overline{k}])$ if they reach at least the welfare level $W_1(\widehat{k}_1, K_1, t)$ $(W_2(\widehat{k}_2, K_2, t))$.

For the case of an exogenous tax, the definition of stability becomes as follows. Given the tax rate $t > 0$, a system of FTS $\mathcal{S}(t) = \left\langle C^1, C^2, \ldots, C^n \right\rangle$ is *stable* if there does not exist an equilibrium for a free trade space T such that for every country of type $i \in T_i$ (for $i = 1,2$):

$$W_i (k (T), K_i, t) > W_i \left(k \left(C^j\right), K_i, t\right) \text{ with } i \in C_i^j,\ C^j \in \mathcal{S}(t).$$

THEOREM 2: There always exist stable systems of free trade spaces for the world economy E.

We have four different types of stable systems of FTS:

1. $\mathcal{S}_1(t) = \langle C^1, C^2 \rangle$ with $\lambda(C^1) = c^1 < e_1$,

$$k(C^1) = K_1 \text{ and } \lambda(C^2) = (e_1 - c^1) + e_2,\ k(C^2) = k_2^{\max}.$$

In this stable system of FTS, all countries of type 2 cooperate with some countries of type 1 in order to achieve their maximum welfare level, leaving the other countries of type 1 alone (as in Figure 2).

2. $\mathcal{S}_2(t) = \langle C^1, C^2 \rangle$ with $\lambda(C^1) = c^2 < e_2$,

$$k(C^1) = K_2 \text{ and } \lambda(C^2) = (e_2 - c^2) + e_1,\ k(C^2) = k_1^{\max}.$$

In this stable system of FTS, all countries of type 1 cooperate with some countries of type 2 in order to achieve their maximum welfare level, leaving the other countries of type 2 alone (as in Figure 3).

3. $\mathcal{S}_3(t) = \langle E \rangle$ with $k(E) = \overline{k}$. In this stable system of FTS, all the countries in the world economy form a unique FTS (as in Figure 4).

4. $\mathcal{S}_4(t) = \langle C^1, C^2 \rangle$ with $\lambda(C^1) = \lambda(C_1^1) + \lambda(C_2^1)$, $k(C^1) = \widehat{k}_1 < \overline{k}$, and $\lambda(C^2) = (e_1 - \lambda(C_1^1)) + (e_2 - \lambda(C_2^1))$, $k(C^2) = \widehat{k}_2 > \overline{k}$. In this stable system of FTS, there exist two FTS formed by both types of countries such that countries of type 1 (type 2) obtain at least the welfare level corresponding to their best FTS in the interval $[K_1, \overline{k}]$ $([\overline{k}, K_2])$ (as in Figure 5).

It is well known in the international trade theory that, under free trade and in the absence of externalities, the total cooperation between countries of different type is a Pareto optimum situation that maximizes the world welfare level. But, in the presence of an externality, Theorem 2 asserts that, under certain conditions, the cooperation between countries of different type is stable when a GATT arbitrator is involved. The resulting stable system of free trade spaces is not necessarily the total cooperation between countries, and it will depend on the tax production rate set by the GATT arbitrator.

We have shown that in Example 1, without the intervention of the GATT arbitrator, the unique stable system of FTS is such that no cooperation occurs between countries of different type. But, if the GATT arbitrator applies a tax production rate $t = 0.1$ we have $\overline{K}(t) = \{K_1\} \cup [41.2, K_2]$, and the unique stable system of FTS is such that partial cooperation occurs: $\mathcal{S}(t) = \langle C^1, C^2 \rangle$ with $k(C^1) = K_1$ and $k(C^2) = 41.2$. However, if the GATT arbitrator applies a tax production rate of $t = 0.2$ we have $\overline{K}(t) = \{K_1\} \cup [21.8, K_2]$ with $\overline{k} = 30 \in \overline{K}(t)$. In this case, the stable system of FTS is such that all countries of the world cooperate: $\mathcal{S}(t) = \langle E \rangle$ with $k(E) = \overline{k} = 30$.

4.4 **An Optimal Tax Policy Based on the Polluter Pays Principle**

Given the intervention margin of the GATT arbitrator, one natural objective is to maximize the world welfare. Not always it is possible to provide a specific rule for the action of the GATT arbitrator. However, under the following assumption, we can design a policy for the GATT arbitrator that maximizes the world welfare.

ASSUMPTION 5: There exists $\underline{K}, \overline{K} \in \mathbb{R}_+$ such that the indirect welfare function of any country of type i being alone (or not cooperating with countries of different type), $W_i(K_i) = f(K_i) - \underline{v}(f(K_i))$, is concave for all $K_i \in [\underline{K}, \overline{K}]$, $i = 1,2$, and $K_1, K_2 \in [\underline{K}, \overline{K}]$.

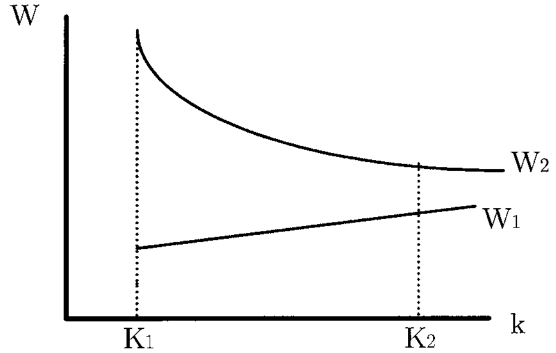
Since the production function $\underline{f}(K_i)$ is concave, assuming the concavity of $W_i(K_i)$ within the interval $[\underline{K}, \overline{K}]$ simply means that the biggest the capital endowment the biggest the external effect.

Next, Theorem 3 shows that the GATT arbitrator can reach the stability of the FTS in which all countries in the world economy cooperate. In this case, the tax production rate is determined, according to the polluter pays principle, by the marginal disutility of each possible capital-labor ratio $k \in (K_1, K_2)$. Moreover, such tax policy maximizes the world welfare level. But, in order to adopt this optimal tax policy, the GATT arbitrator needs to know the welfare functions of the countries.

THEOREM 3: Let $t = v'(f(k(C)))$ be the tax production rate chosen by the GATT arbitrator for any FTS C with capital-labor ratio $k(C) \in (K_1, K_2)$. Then, there exists a unique stable system of FTS $\mathcal{S}(t) = \langle E \rangle$ and it maximizes the world welfare level.

When the tax production rate is established according to the marginal disutility caused by the externality in each FTS, the indirect welfare functions of both types of countries are of the form represented in Figure 6

FIGURE 6



In this case, both types of countries would like to cooperate in the formation of a FTS with countries of different type. This result follows from the fact that there is no market failure, given that pollution is confined to the country of origin and that pollution is regulated optimally by national governments. Hence, the unique stable system of FTS is such that all countries of the world economy are cooperating in a unique FTS. Moreover, Assumption 5 guarantees us that this stable system of FTS maximizes the world welfare level.

5 Conclusions

Using a cooperative game-theoretic approach we have studied the impact of a depletable externality on the formation and stability of free trade agreements in a general equilibrium framework. We have shown that there exist five different equilibrium structures of free trade spaces that could prevail according to the notion of the core. In order to avoid the Pareto dominated equilibrium structure sovereign countries should regulate the depletable externality.

The main contribution of the paper is methodological: it provides a framework for studying the role of private externalities on the formation and stability of free trade spaces. Its main message is that in order to establish a globally Pareto optimal trading system, it may be necessary to tax those that have much to gain from it and subsidize those that do not. While perhaps unappealing at first glance, the transfers implemented by means of the tax production rate have the attractive feature of inducing countries with large positive externalities on other countries to take part in the global trading process. In a world where situations of « *laissez-faire* » fail to be Pareto optimal, we advocate in favor of an IEA or the set-up of a GATT arbitrator in order to reach a Pareto improvement in the presence of a private externality.

In our simple model with a unique traded good and starting in autarchy, the formation of any partial or total free trade space between countries of different type is always welfare improving for both types of countries (in the absence of the negative externality). Following GROSSMAN [1984] and SVENSSON [1984], this result is due to the fact that the initial situation is zero trade (there is no terms-of-trade effect due to the introduction of factor trade) and that the final situation is free trade (there is no volume-of-trade effect when factor trade is introduced).[14]

In our model, trade integration between countries of different type is always beneficial to the capital rich countries and ambiguous for the capital poor countries. This result is quite sensitive to the fact that there is only one good produced in the economy. In a Heckscher-Ohlin model with two goods

14. This is also the main message of KEMP and WAN [1976]. They showed that, in a competitive world economy in which transfers and non-optimal tariffs are allowed, there always exists a transfer scheme and a common tariff vector that makes enlarging a customs union better for all member countries while non-members remain indifferent.

produced the result could be inverted. Assuming, for example, that capital rich countries have a comparative advantage in the polluting sector, by getting more specialized in this sector the rich economies would have to trade off traditional gains from trade against increased local negative externality.

The trend in neoclassical development establishes that foreign direct investment of transnational corporations, left to free market forces, benefits host countries: it increases local skills, technology and production, it provides competition for local firms, it creates additional exports, and some of the skills and knowledge they create may spill over to the rest of the economy (see, LALL [1998]). For such positive externalities, the results would be inverted: capital poor countries would benefit from free trade but trade would have an ambiguous effect for capital rich countries.

Some extensions may be worthwhile. In order to avoid the existence of strategic externalities between free trade spaces, we have assumed that countries participating in a free trade space set a common external tariff sufficiently high so as to induce autarchy for each space. Under this assumption, the cooperative game-theoretic approach applies provided that welfare of each free trade space is independent of the coalition structure formed. Hence, a first extension is the question of how external tariffs might be determined simultaneously with the choice of free trade space. To answer this question one could adopt, as in YI [1996], recent developments in the non-cooperative theories of stable coalition structures (BLOCH [1996] and YI [1995]).[15]

Finally, we have assumed that countries differ only in their capital endowments but that all countries in the world economy have the same welfare function (or the same environmental standard). A second extension is to introduce countries with different environmental standards. In this case, one should enlarge the set of instruments available to regulate the negative externality.

• References

BAUMOL W.J., OATES W.E. (1988). – *The Theory of Environmental Policy*, Cambridge University Press, Cambridge, UK.

BLOCH F. (1996). – « Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division », *Games and Economic Behavior*, 14, pp. 90-123.

BOND E.W., SYROPOULOS C. (1996). – « The Size of Trading Blocs, Market Power and World Welfare Effects », *Journal of International Economics* 40, pp. 411-437.

COPELAND B.R., TAYLOR M.S. (1994). – « North-South Trade and The Environment », *Quarterly Journal of Economics*, 109, pp. 755-787.

COPELAND B.R., TAYLOR M.S. (1995). – « Trade and Transboundary Pollution », *American Economic Review*, 85, pp. 716-737.

DAGAN N., VOLIJ O. (1997). – « Formation of Nations in a Welfare-State Minded World », *Mimeo*, Department of Economics, Brown University, Providence, Rhode Island.

DIXIT A.K., NORMAN V. (1980). – *Theory of International Trade*, Cambridge Economic Handbooks University Press, Cambridge, UK.

15. YI [1996], in a model where the formation of customs unions reduces the welfare of non-member countries, has shown that global free trade is an equilibrium outcome under the open regionalism rule but not under the unanimous regionalism rule.

FREEMAN A.M. (1984). – « Depletable Externalities and Pigouvian Taxation », *Journal of Environmental Economics and Management*, XI, pp. 173-179.

GRAFE F., MAULEON A. (1997). – « Externalities and Free Trade Agreements » S.E.E.D.S. *Discussion Paper* 172.

GROSSMAN G.M. (1984). – « The Gains from International Factor Movements », *Journal of International Economics*, 17, pp. 73-83.

KEMP M.C., WAN H. (1976). – « An Elementary Proposition Concerning the Formation of Customs Unions », *Journal of International Economics*, 6, pp. 95-97.

KENNAN J., RIEZMAN R. (1990). – « Optimal Tariff Equilibria with Customs Unions », *Canadian Journal of Economics*, 23, pp. 70-83.

KOWALCZYK C., SJOSTROM T. (1994). – « Bringing GATT into the Core » *Economica*, 61, pp. 301-317.

KRUGMAN P.R. (1991). – « Is Bilateralism Bad? » in Elhanan Helpman and Assaf Razin (eds.), *International Trade and Trade Policy*, Cambridge, Mass.: MIT Press.

KRUGMAN P.R., OBSTFELD M. (1991). – *International Economics: Theory and Policy*, 2nd Ed., Harper Collins Publishers, New York.

LALL S. (1998). – « Changing Perceptions of Foreign Direct Investment in Development » in *International Trade, Foreign Direct Investment and the Economic Environment, Essays in Honour of Professor Sylvain Plasschaert*, edited by P.K.M. Tharakan and D. Van Den Bulcke, Macmillan Press Ltd, Great Britain.

LEVY P.I. (1997). – « A Political-Economic Analysis of Free Trade Agreements », *American Economic Review*, 87, pp. 506-519.

LOW P., YEATS A. (1992). – « Do Dirty Industries Migrate? » in Patrick Low, ed., *International Trade and the Environment*; World Bank *Discussion Papers* (Washington, DC: World Bank, 1992).

LUCAS R.E.B., WHEELER D., HETTIGE H. (1992). – « Economic Development, Environmental Regulation and the International Migration of Toxic Industrial Pollution: 1960-1988 » in Patrick Low, ed., *International Trade and the Environment*; World Bank *Discussion Papers* (Washington, DC: World Bank, 1992).

MACHO-STADLER I., PEREZ-CASTRILLO D., PONSATI C. (1998). – « Stable Multilateral Trade Agreements », *Economica*, 65, pp. 161-177.

MALER K.G. (1990). – « International Environmental Problems », *Oxford Review of Economic Policy*, 6(1), pp. 80-108.

MARKUSEN J.R. (1975). – « International Externalities and Optimal Tax Structures », *Journal of International Economics*, 5, pp. 15-29.

PEARCE D. (1995). – *Capturing Global Environmental Value*, London: Earthscan Publications Limited.

PETHIG R. (1976). – « Pollution, Welfare and Environmental Policy in the Theory of Comparative Advantage », *Journal of Environmental Economics and Management*, II, pp. 160-169.

RIEZMAN R. (1985). – « Customs Unions and the Core », *Journal of International Economics*, 19, pp. 355-365.

SIEBERT H.,. EICHBERGER J, GRONYCH R., PETHIG R. (1980). – *Trade and Environment: A Theoretical Enquiry*, Amsterdam: Elsevier/North-Holland, 1980.

SVENSSON L.E.O. (1984). – « Factor Trade and Goods Trade », *Journal of International Economics*, 16, pp. 365-378.

YI S-S. (1995). – « Stable Coalition Structures with Negative External Effects », *Working Paper* 95-9, Dartmouth College.

YI S-S. (1996). – « Endogenous Formation of Customs Unions under Imperfect Competition: Open Regionalism is Good », *Journal of International Economics*, 41, pp. 153-177.

APPENDIX

In order to prove Theorem 1 a lemma similar to Lemma 2 in DAGAN and VOLIJ [1997] is needed. The reader is referred for a proof to DAGAN and VOLIJ [1997].

LEMMA 1. Let C be a FTS with capital-labor ratio $k(C) = [c_1K_1 + c_2K_2] / [c_1 + c_2]$. Let k' and k'' be two arbitrary numbers satisfying

$$K_1 \leqslant k' < k(C) < k'' \leqslant K_2.$$

Then, there exists a unique partition of the free trade space C into two free trade spaces, T and R, such that $C = T \cup R$ and $k(T) = k'$, $k(R) = k''$.

Lemma 1 says that if there exists a FTS C with countries of both types, it is always feasible to form two new free trade spaces by choosing suitable proportions of countries of each type, with capital-labor ratios equal to any two numbers, k' and k'', satisfying $K_1 \leqslant k' < k(C) < k'' \leqslant K_2$.

Proof of Theorem 1

Taking into account that countries of type 2 would like to participate in any cooperation agreement with countries of type 1, the stable systems of free trade spaces will be determined by the characteristics of the indirect welfare function of countries of type 1. The proof is divided in two main cases.

1. $\overline{K} = \{K_1\}$. Then, there exists a unique stable system of FTS $\mathcal{S}_0 = \left\langle C^1, C^2 \right\rangle$ with $k\left(C^1\right) = K_1$ and $k\left(C^2\right) = K_2$. $\mathcal{S}_0$ is stable because any other FTS implies a lower welfare for countries of type 1, and is unique since we consider only systems of FTS in which all FTS have different capital-labor ratio.

2. $\overline{K} \neq \{K_1\}$. We distinguish two different cases depending $K_1^{\max}$ is or is not a singleton. We present the proof for the case $K_1^{\max}$ is a singleton: $K_1^{\max} = \{k_1^{\max}\}$. The reader is referred to GRAFE and MAULEON [1997] for the case where $K_1^{\max}$ is not a singleton.

2.a. Consider the case in which $K_1^{\max}$ is a singleton: $K_1^{\max} = \{k_1^{\max}\}$. We divide case **2.a** in three subcases.

2.a.1. $\bar{k} < k,\ \forall k \in \overline{K},\ k \neq K_1$. There exists a unique stable system of FTS $\mathcal{S}_1 = \left\langle C^1, C^2 \right\rangle$ with $k\left(C^1\right) = K_1$

and $$k(C^2) = \arg\max_{k \in \overline{K}} W_2\left(k(C), K_2\right) = \min\left\{k : k \in \overline{K}, k \neq K_1\right\}.$$

$\mathcal{S}_1 = \langle C^1, C^2 \rangle$ is stable since any other individually rational FTS

C, $k(C) > k(C^2)$, implies a lower welfare for countries of type 2. $\mathcal{S}_1 = \langle C^1, C^2 \rangle$ is unique. Assume on the contrary that there exists a $\mathcal{S}_1' = \langle C^*, C^{**} \rangle$, $C^* \subset E_1$ and $k(C^{**}) > k(C^2)$. It is easy to see, using Lemma 1 and the continuity of the welfare functions of both types of countries, that $\mathcal{S}_1'$ is not stable since a coalition formed by all countries of type 2 in C^{**} and some positive measure of countries of type 1 not in C^{**} would be better off forming a FTS C with $k(C^2) < k(C) < k(C^{**})$, $k(C) \in \overline{K}$.

2.a.2. $\overline{k} > k$, $\forall k \in \overline{K}$. Then, there exists a unique stable system of FTS $\mathcal{S}_2 = \langle C^1, C^2 \rangle$ with $k(C^1) = K_2$ and $k(C^2) = k_1^{\max}$. $\mathcal{S}_2 = \langle C^1, C^2 \rangle$ is stable since any other individually rational FTS C, $k(C) \neq k_1^{\max}$, implies a lower welfare for countries of type 1. $\mathcal{S}_2 = \langle C^1, C^2 \rangle$ is unique. Assume on the contrary that there exists a $\mathcal{S}_2' = \langle C^*, C^{**} \rangle$, $C^* \subset E_2$ and $k(C^{**}) \neq k(C^2)$. Two cases arise:

i) If $\overline{k} > k(C^{**}) > k_1^{\max}$, $\mathcal{S}_2'$ is not stable since a coalition formed by all countries of type 1 and some positive measure of countries of type 2 in C^{**}, improve their welfare if they form the FTS C^2.

ii) If $\overline{k} > k_1^{\max} > k(C^{**})$, $\mathcal{S}_2'$ is not stable since a coalition formed by all countries of type 1 and some positive measure of countries of type 2 not in C^{**}, improve their welfare if they form the FTS C^2.

2.a.3. $\min\{k : k \in \overline{K}, k \neq K_1\} < \overline{k} < \max\{k : k \in \overline{K}\}$. Three sub-cases arise.

2.a.3.1. $k_1^{\max} < \overline{k}$. There exists a unique stable system of FTS $\mathcal{S}_2 = \langle C^1, C^2 \rangle$ with $k(C^1) = K_2$ and $k(C^2) = k_1^{\max}$. The proof is the same than in **2.a.2**.

2.a.3.2. $k_1^{\max} > \overline{k}$ and $\arg\max_{k \in \overline{K}, k \leqslant \overline{k}} W_1(k(C), K_1) = \overline{k}$. There exists a unique stable system of FTS $\mathcal{S}_3 = \langle E \rangle$ with $k(E) = \overline{k}$. $\mathcal{S}_3 = \langle E \rangle$ is stable because any potential FTS that countries of type 1 would like to form, needs some positive measure of countries of type 2, but no positive measure of countries of type 2 agree to participate in such FTS. The same argument applies in any potential agreement that countries of type 2 would like to form. $\mathcal{S}_3 = \langle E \rangle$ is unique.

Consider a system $\mathcal{S}_3' = \langle C^1, C^2 \rangle$ with $k(C^1) < \overline{k} < k(C^2)$.

$\mathcal{S}_3'$ is not stable since a positive measure of countries of type 1 in C^1 and a positive measure of countries of type 2 in C^2 can be better off by forming a FTS C with $k(C) = \overline{k}$.

2.a.3.3. $k_1^{\max} > \overline{k}$ and $\widehat{k} = \arg \max\limits_{k \in \overline{K}, k \leqslant \overline{k}} W_1(k(C), K_1) < \overline{k}$. There exists a unique stable system of FTS $\mathcal{S}_4 = \langle C^1, C^2 \rangle$ with $k(C^1) = \widehat{k}$ and

$$k(C^2) = k^* = \underset{k \in \left[k | k \geqslant \overline{k}, W_1(k(C), K_1) = W_1\left(\widehat{k}, K_1\right)\right]}{argmax} W_2(k(C), K_2).$$

$\mathcal{S}_4 = \langle C^1, C^2 \rangle$ is stable because countries of type 1 would like to form a FTS C with $k(C) > k^*$, but no positive measure of countries of type 2 would like to enter into such space. Moreover, the positive measure of countries of type 2 in C^2 would like to form a FTS C with $k(C) < k^*$, but no positive measure of countries of type 1 would like to participate in such FTS. $\mathcal{S}_4 = \langle C^1, C^2 \rangle$ is unique. Assume on the contrary that $\exists C^*$ with $k(C^*) \neq \widehat{k}, k^*$, $C^* \in \mathcal{S}_4' \neq \mathcal{S}_4$.

i) If $k(C^*) < \widehat{k}$, the positive measure of countries of type 1 in C^* and some positive measure of countries of type 2 not in C^* are better off forming the FTS C^1.

ii) If $\widehat{k} < k(C^*) < k^*$, the positive measure of countries of type 1 in C^* and some positive measure of countries of type 2 in C^* are better off forming the FTS C^1.

iii) If $k(C^*) > k^*$, the positive measure of countries of type 2 in C^* and some positive measure of countries of type 1 not in C^* (alternatively, a subgroup of countries of type 1 in C^*) are better off forming a FTS C with $k^* \leqslant k(C) < k(C^*)$. ■

Proof of Proposition 1

1. $\overline{K} = \{K_1\}$. We have $\mathcal{S}_0 = \langle C^1, C^2 \rangle$ with $C^1 = (E_1, \emptyset)$ and $C^2 = (\emptyset, E_2)$. Then, $V(\mathcal{S}_0) = e_1 W_1(K_1) + e_2 W_2(K_2)$.

2. $\overline{K} \neq \{K_1\}$. Remember that we have divided this case in three subcases (see the graphic explanation of Theorem 1).

2.a. $\mathcal{S}_1 = \langle C^1, C^2 \rangle$ is the stable system of FTS where $\lambda(C^1) = c^1 < e_1$ with $k(C^1) = K_1$ and $\lambda(C^2) = (e_1 - c^1) + e_2$ with

$$k(C^2) = k^* = \arg \max_{k \in \overline{K}} W_2(k(C), K_2).$$

Then, $V(\mathcal{S}_1) = c^1 W_1(K_1) + (e_1 - c^1) W_1(k^*, K_1) + e_2 W_2(k^*, K_2)$. Since $W_1(k^*, K_1) \geqslant W_1(K_1)$ and $W_2(k^*, K_2) > W_2(K_2)$, we have $V(\mathcal{S}_1) > V(\mathcal{S}_0)$.

2.b. $\mathcal{S}_2 = \langle C^1, C^2 \rangle$ is the stable system of FTS where $\lambda(C^1) = c^1 < e_2$ with $k(C^1) = K_2$, $\lambda(C^2) = (e_2 - c^1) + e_1$ with

$$k(C^2) = k_1^{\max} = \arg\max_{k \in \overline{K}} W_1(k(C), K_1) < \overline{k}.$$

Then, $V(\mathcal{S}_2) = c^1 W_2(K_2) + (e_2 - c^1) W_2(k_1^{\max}, K_2) + e_1 W_1(k_1^{\max}, K_1)$. Since $W_2(k_1^{\max}, K_2) > W_2(K_2)$, $W_1(k_1^{\max}, K_1) \geqslant W_1(K_1)$, we have $V(\mathcal{S}_2) > V(\mathcal{S}_0)$.

2.c. Three different classes have been distinguished (see the graphic explanation of Theorem 1). The first one reverts to (2.b). In the second class, the stable system of FTS is $\mathcal{S}_3 = \langle E \rangle$ where $k(E) = \overline{k}$. Then, $V(\mathcal{S}_3) = e_2 W_2(\overline{k}, K_2) + e_1 W_1(\overline{k}, K_1)$. Since $W_2(\overline{k}, K_2) > W_2(K_2)$ and $W_1(\overline{k}, K_1) \geqslant W_1(K_1)$, we have $V(\mathcal{S}_3) > V(\mathcal{S}_0)$. In the third class, the stable system of FTS is $\mathcal{S}_4 = \langle C^1, C^2 \rangle$ where $\lambda(C^1) = \lambda(C_1^1) + \lambda(C_2^1)$ with $k(C^1) = k^* < \overline{k}$, and

$$\lambda(C^2) = (e_1 - \lambda(C_1^1)) + (e_2 - \lambda(C_2^1)) \text{ with } k(C^2) = k^{**} > \overline{k}.$$

Since $W_2(k^*, K_2) > W_2(K_2)$, $W_2(k^{**}, K_2) > W_2(K_2)$, and $W_1(k^*, K_1) = W_1(k^{**}, K_1) \geqslant W_1(K_1)$, we have

$$V(\mathcal{S}_4) = \lambda(C_2^1) W_2(k^*, K_2) + (e_2 - \lambda(C_2^1)) W_2(k^{**}, K_2) + e_1 W_1(k^*, K_1) > V(\mathcal{S}_0). \quad \blacksquare$$

Proof of Theorem 2

See, GRAFE and MAULEON [1997].

Proof of Theorem 3

Firstly, we show that, given any tax production rate, the world welfare level is maximized when all countries cooperate. Only four types of stable systems of FTS are possible (see the explanation after Theorem 2). Next we compare the welfare levels associated to these four types, where

$$V(\mathcal{S}_1(t)) = ((e_1 - c^1) + e_2)[f(k_2^{\max}) - v(f(k_2^{\max}))] + c^1[f(K_1) - v(f(K_1))];$$

$$V(\mathcal{S}_2(t)) = \left(e_1 + \left(e_2 - c^2\right)\right)[f(k_1^{\max}) - v(f(k_1^{\max}))]$$
$$+c^2[f(K_2) - v(f(K_2))];$$

$$V(\mathcal{S}_3(t)) = (e_1 + e_2)[f(\overline{k}) - v(f(\overline{k}))];$$

$$V(\mathcal{S}_4(t)) = (\lambda(C_1^1) + \lambda(C_2^1))[f(\widehat{k_1}) - v(f(\widehat{k_1}))]$$
$$+[(e_1 - \lambda(C_1^1)) + (e_2 - \lambda(C_2^1))][f(\widehat{k_2}) - v(f(\widehat{k_2}))].$$

From Assumption 5, it is straightforward that $V(\mathcal{S}_3(t)) \geqslant V(\mathcal{S}_l(t))$, $l = 1,2,4$.

Secondly, we show that, given the tax production rate $t = v'(f(k(C)))$, the FTS $\mathcal{S}(t) = \langle E \rangle$ is the unique stable system. The indirect welfare function of any country of type $i \in C_i$, $i = 1,2$, in the FTS C with the tax production rate t is:

$$W_i(k(C), K_i, t) = f(k(C)) - (1-t) f'(k(C))(k(C) - K_i)$$
$$-v(f(k(C))),$$

with $\partial W_i / \partial k(C) = t f'(k(C)) - (1-t) f''(k(C))(k(C) - K_i)$
$$-v'(f(k(C))) f'(k(C)).$$

Substituting $t = v'(f(k(C)))$, we have

$$\partial W_i / \partial k(C) = -\left(1 - v'(f(k(C)))\right) f''(k(C))(k(C) - K_i)$$

which is positive for countries in E_1 and negative for countries in E_2. Therefore, if the tax production rate is equal to the marginal disutility caused by the externality in each FTS C [with $k(C) \in (K_1, K_2)$], then the indirect welfare functions of both types of countries are of the form represented in Figure 6 and the FTS in which all countries of the world economy are cooperating is stable. ■