

RTAs formation and trade policy*

Davide Sala
European University Institute

First Draft: June 2005
This draft: September 2006

Abstract

From the mid-60s to the mid-80s there has been a gradual but fundamental change in the nature of trade protection. International trade has become increasingly restricted by quotas and other nontariff barriers, as the level of tariffs have fallen and governments have devised other forms of protection for sectors facing increased foreign competition.

The paper shows such non-tariff barriers have very different effects and implications from tariff for the welfare outcome of a regional integration agreement. Indeed, binding quotas, differently from tariffs, succeed to preserve the trade volumes with the rest of the world, and lead to welfare improving customs unions and free trade areas since trade between the partners is not expanded at the expense of trade with the outside world. By relating the existence of welfare enhancing regional integration to the systematic change in the type of trade policy conducted by most countries, this paper emphasizes that the desirability of piecemeal reforms has increased through time and justifies a renewed and grown policy interest in preferential trade in the 90s, when NTBs had a greater weight in trade policies. This can contribute to explain the spurt of regionalism observed in the data.

* I am indebted to Prof. Staiger Robert for his useful suggestions and comments. I also would like to thank Prof. Omar Licandro, Prof. Morten Ravn, Philip Sauré and all other participants to the seminar at EUI. All mistakes are mine.

* Author's corresponding address is davide.sala@iue.it

1 Introduction

Starting from the 90s, regional integration has experienced a spurt and it has represented a major feature of International Relations. Starting in 1948, some 200 RIAs have been notified to GATT/WTO, but about one-half of all notifications have occurred in the last decade.

This recent proliferation of preferential trade agreements (PTAs) was often attributed to the unsatisfactory progress of multilateral trade negotiations and therefore, to issues related to the efficiency of multilateral trade liberalization to achieve free trade.¹ This view has shifted the economic debate from the traditional concern about the welfare implications of preferential trade *per se*, to the highly controversial issue on whether regionalism or multilateralism is the most effective strategy for achieving global free trade.²

A common feature of this literature is that trade policy has been represented mostly by tariff measures and their reductions, mirroring probably that the major effort of GATT rounds was the reduction of tariff barriers.³

However, this is at odds with the striking empirical observation that numerous other in-

¹ See for example Bhagwati (1993). The first wave of regionalism took place in the 1960s, but it failed to spread because the U.S. supported a multilateral approach. But the U.S. changed positions, and – starting with the 1980s – has favored regional trade agreements. See also Fernandez (1997) for a broad and extensive review on the "traditional" and "non-traditional" gains of regionalism .

² Some view PTAs as antithetical to the GATT, inevitably leading to a world of warring trade blocs - *stumbling blocs* in Bhagwati's definition - while others view them as supplemental, being just one more path by which global free trade can be approached - *building blocs*. See Summers (1991), Krugman (1993) and Frankel et. al (1997). For a traditional approach to custom unions (CUs) focusing on its welfare outcome, refer to Viner (1950).

³ In this literature, alternative measures to tariff protection such as quota restrictions have been analyzed in terms of their tariff-equivalent. Note that uniform tariff measures are also analytically convenient.

struments of trade policy - such as global and country specific volume quotas, value quotas, voluntary export restraint (VER), local content schemes - have become more and more popular throughout all countries. The relevance of this is that Corden and Falvey (1985) and Falvey (1988) have shown that these instruments have specific implications - not shared with tariffs - in the context of piecemeal tariff reform which are likely to be relevant for regional integration too, but have not been - to the best of my knowledge - explored so far. Going back to a more traditional welfare approach, the aim of this paper is to show that, differently from tariffs, such non tariff barriers have positive implications for the welfare consequence of a regional integration agreement.

Starting from this standpoint, this paper relates the policy interest in preferential trade among nations to a well known feature of international trade relations - the gradual but fundamental change in the nature of trade protection over the past thirty years. As international negotiation has succeeded in reducing tariffs to low levels, national governments have resorted to a range of increasingly intricate policies to protect their domestic industries from foreign competition. As Laird and Yeats (1990) note "*while a major effort was made in multilateral trade negotiations to reduce tariffs, protectionism in the form of NTBs (non tariff barriers) greatly expanded, and may have even offset or exceeded the effects of liberalized import duties*". Estimates relating to the post WWII period indicate that the average tariff on manufactures in developed countries was approximately 40 percent, whereas nowadays it is lower than 5 percent. The declining proportion of imported industrial products that do not receive duty free treatment in developed countries, and the reduced fraction of imports

into developed countries from all sources facing tariff rates more than 15%, point to the same evidence of falling tariffs protection over the same period. On the contrary, in 1966 NTBs affected 25 per cent of developed countries' imports, while in 1986 this share had increased to 48 percent, a symptom of an increasing trend of non-tariff trade restrictions.

In spite of GATT prohibitions, these quantitative restriction have been widely used, partly because GATT rules do permit countries to recourse to instruments such as quota to protect domestic agriculture as anti-dumping measures and as temporary "safeguard" actions. Moreover, most developing economies are exonerated from these rules and trade in textile and apparel and agriculture commodities has been largely exempted from GATT disciplines for decades. In addition, VER, being voluntary, are negotiated on a bilateral basis outside the GATT framework.⁴ The consequence being that such direct quantitative restrictions (QRs) on international trade have greatly expanded their coverage: from \$ 30 billion of imports from OECD countries affected by NTBs in 1966 (\$ 100 billion in 1986 prices) to the \$356 billion of 1986.

In the context of regional integration (RI), non tariff barriers have different implications from tariffs on trade flows and therefore, they lead to a different welfare outcome. Prior to a regional agreement, it may be presumed that a country had the same tariffs on all countries, but it purchased from outside the union because that price was lowest. After the union, the

⁴ For instance, The Long-Term Textile Agreement (LTA) and the Multi-Fiber Arrangement (MFA) are a series of bilateral trade quotas covering textiles and apparel that have been negotiated outside the GATT. Many countries have kept their programs of support for agricultural apart from GATT rules. The USA, for example, have being using import quotas to support domestic farm price, and similarly the European Community supports its farmers with its Common Agricultural Policy which relies on a series of variable tariffs that act very much like quotas.

country switches its purchases from the lowest-price to a higher-price country. However, if imports had been limited by volume quotas, a member-country could still have imported the good from the lowest cost supplier, as long as the quantity of imports did not exceed the quota limits. Trade diversion occurs at a less extent under quota restricted trade than under tariff-restricted trade. In particular, if quotas vis-à-vis countries outside the union result binding even after the implementation of a RTA, the volumes of trade with the rest of the world will be clearly preserved - which is the "*efficient rule*" for welfare-enhancing custom unions and FTAs as proposed by Kemp and Wan (1976) and Krishna and Panagariya (2002). In other words, the Kemp-Wan and Krishna-Panagariya results have important implications in the presence of QRs. This is the main result of the paper and suggests welfare gains from a trade agreement may turn to be more likely under NTBs than under tariff-barriers. In turn, this may have led government to a renewed policy interest in the formation of RTAs after 1990.

The intuition for this result is simple: binding quotas, differently from tariffs, succeed to preserve the trade volumes with the rest of the world. When purchases from the rest of the world are fixed, countries in a CU and in a FTA move closer to free trade which is welfare improvement. Indeed, in this specific case the FTA does not reduce trade with the outside world and expands trade between the partners.

Krishna and Panagariya (2002) do not provide a general existence result for a tariff vector ensuring that the criterion they propose for welfare enhancing FTA holds - namely, every FTA member imports and exports the same quantity with the rest of the world as they did

before the FTA. This paper shows that the presence of binding NTBs vis-à-vis non member countries leads to a FTA formation that naturally implements this "efficiency rule", whereas the same agreement implemented in presence of tariff restricted trade could not lead to such outcome. The intuition for the latter result is related to Bagwell and Staiger (2004) who shows that any bilateral agreement on tariff-reduction can preserve the pre-existing volumes of trade if only satisfies the principles of reciprocity and Most Favoured Nation (MFN) treatment, which is necessarily violated by a FTA.⁵

Furthermore, Kemp and Wan (1976) and Krishna and Panagariya (2002) allow us to conclude that welfare improving CUs and FTAs exist, while this paper emphasizes that the desirability of such reforms has increased through time, by relating the existence of welfare enhancing FTAs to the systematic change in the type of trade policy instruments adopted by most countries over the past twenty-five years. As countries have reduced tariff protection in the fulfilment of GATT regulations and substitute it with non-tariff trade restrictions, policy maker's interest for joining into a RTAs has also increased, resulting in a greater formation of RTAs.

Chang and Winters (2002) have shown that Mercosur has brought significant terms of trade appreciation for Brazil and loss for the external countries exporting into the area, including US, Japan, Germany, South Korea and Chile. Interestingly, the model shows that

⁵ This principle states that all countries belonging to GATT should be treated equally. So if Japan reduces its tariffs on goods coming from Europe, it must do the same for those goods coming from any other member country of GATT.

Under the principle of reciprocity, negotiations result in tariff adjustments that generate for each participant an equal change in the volume of its imports and exports.

such effects could be driven by agreements upon the abolition of existing VERs (possibly arising also in the context of regional integration) and that such agreements may turn to be welfare-enhancing. This is especially important, given the extensive use of VERs in the mid-80s for which later on countries have rescinded their requests in the 90s.⁶

Finally, in their survey De Melo et. al. (1993), they briefly discuss the possibility of a welfare improving FTA in presence of QRs, but their underline assumption is that partner countries are small relative to the rest of the world. I consider the large country setting, but like them, I assess the benefits of RI from the view point of participating countries rather than the world as a whole.

The basic insights are developed in a three-countries, two-goods general equilibrium trade model presented in section 2. I shall follow closely Bagwell and Staiger (1999), (2002) and (2004), and extend their framework to include global quotas, country-specific quotas and VERs besides tariffs, as instruments of trade policy available to countries. Then, the model is used in section 3 to analyze the different welfare consequence of a RTA in presence of quota-restricted from tariff-restricted trade. The small open economy case is illustrative of the different adjustment mechanisms implied by NTBs and tariffs after the implementation of an agreement. Such mechanisms will still be operative in the large economy case, but they will be either reinforced or offset by terms of trade effects. Given the empirical relevance of VERs, section 4 shows that they can also lead to a favorable welfare outcome of regionalism in spite of the terms of trade effects induced. In section (5), I turn to the possibility of

⁶ For example Laird and Yeat (1990) have documented that the incidence of VERs on import volumes has greatly increased between 1981 and 1986 for all developed economies.

comparing the gains arising from RIAs under two different trade policies, namely quotas and tariffs. The last section concludes.

2 The model

The basic insights for the different outcome of regional integration in presence of two different instruments of trade policy - namely tariffs and volume quota restrictions - can be developed in a three country $\{A, B, C\}$, two goods $\{x, y\}$ general equilibrium model.⁷ All countries share the same preferences over x, y which are normal goods in consumption and are produced in each country with a non-increasing return to scale technology. All markets are perfectly competitive.

Trade-Patterns

Trade patterns reflect comparative advantages: without loss of generality, country A (home country- no $*$) exports good y to foreign countries $*B$ and $*C$ and imports good x from them. For simplicity, suppose further that the two foreign countries do not trade with one another⁸. This is the easiest and minimal structure to confront, in the next section, the home country with the option of a trade agreement with $*B$ (e.g. a FTA) when trade is restricted either by NTBs or by tariff measures.

Trade Policy

The trade policy conducted by each country is exogenously given, as its political deter-

⁷ This model is developed in greater detail in Bagwell and Staiger (1999), (2002) and (2004).

⁸ This is the simplest way to ensure A can set a discriminatory trade policy (e.g. a trade agreement) against its two foreign trading partners without prohibiting trade with the less favored partner.

mination is beyond the aim of the present paper. Among several types of NTBs, direct quantitative restrictions on international trade have become particularly widespread. Thus, I shall restrict my attention here to volume import quotas and VERs. The volume quota imposed by the home country on imports of x from foreign country i is denoted as \bar{m}^i , and \bar{m}^{*i} is likewise the quantity restriction imposed by foreign country i on imports of good y . Note that, more broadly, \bar{m}^i may be re-interpreted as the VER imposed (possibly on request or threat) by foreign country i to A , and likewise, \bar{m}^{*i} may be the VER levied by A against foreign country i .

When trade protection is achieved through tariffs, the ad valorem tariff that the home government places on imports of x from foreign country i is denoted as t^i , and t^{*i} likewise denotes the ad valorem tariff levied by the government of foreign country i on imports of y from the home country. When $t^B = t^C$, A is implementing a non-discriminatory trade policy complying with the Most Favoured Nation (MFN) principle of GATT. On the other hand, $t^B \neq t^C$ is a discriminatory trade policy. Note that t (t^{*i}) can also be the implicit trade tax associated with a quantitative restriction \bar{m}^i (\bar{m}^{*i}).

Assume that the revenue collected by each country from trade protection is disbursed to the consumers by costless non-distortionary means. Throughout, I shall assume the market for import quota licenses is competitive. That is, quota are allocated so as to ensure perfect competition among quota holders.⁹

⁹ This assumption ensures that in equilibrium there will be no unexploited profit opportunities so that each quota-license holder uses entirely its license.

Moreover, as Krueger (1974) notes quota-induced scarcity rents are economically valuable and may push some firms to engage in some "rent seeking" activities such as lobbying, red-tape, and other resource-using activities. I disregard this possibility here and assume that there is no diversion of economic resources from

International Transaction, and Terms of Trade

Let $p \equiv p_x/p_y$ denotes A 's local relative price and similarly $p^{*i} = p_x^{*i}/p_y^{*i}$, the local relative price in foreign country i . Letting $\tau^i \equiv (1 + t^i)$ and $\tau^{*i} = (1 + t^{*i})$, the local prices can be represented in terms of world prices and tariffs by $p = \tau^i p^{wi} \equiv p(\tau^i, p^{wi})$ and $p^{*i} = p^{wi}/\tau^{*i} \equiv p^{*i}(\tau^{*i}, p^{wi})$. $p^{wi} = p_x^{*i}/p_y$, $i = B, C$ is the "world" (i.e. untaxed) relative price or, equivalently, the ratio of exporter/producer prices for trade between the home country and foreign country i . It represents foreign country i 's terms of trade while its inverse is A 's bilateral terms of trade with each trading partner.¹⁰

Foreign country i buys good y internationally from country A at p_y , but its local price is $p_y^{*i} = \tau^{*i} p_y$, whereas the local price of good x coincides with the producer or the exporter price p_x^{*i} , as no export taxes or subsidies are in place. A has two possible source countries from which to buy good x at either p_x^{*B} or p_x^{*C} . It is natural to assume that consumers in the home country are indifferent between an homogenous good x coming from the two locations B and C .¹¹ Due to this perfect substitutability, A will be trading with both partners provided that consumer prices of goods coming from different locations are equalized:

$$p_x = (1 + t^B)p_x^{*B} = (1 + t^C)p_x^{*C} \quad (1)$$

or in terms of relative price:

their most efficient use.

¹⁰The terms of trade of a country is defined as the price of its exports relative to the price of its imports.
¹¹This preference could be represented by $U(C_x^{*i}, C_y^{*i})$ in foreign country i and $U(C_x, C_y)$ in the home country with $C_x = c_{xA} + c_{xB} + c_{xC}$ being a linear subutility function. C_x is a quantity index giving the total amount of good x consumed regardless of its origin. c_{xk} , $k = A, B, C$ is the units of good x manufactured in country k and consumed in country A . The linearity makes good xk from the different source countries perfect substitutes, or equivalently, the indifference curves between these locations are linear.

$$p = \tau^B p^{wB} = \tau^C p^{wC} \quad (2)$$

which states that world prices are linked across bilateral trading relationships. (2) implies that under MFN tariffs ($\tau^B = \tau^C$) a single world price $p^w = p^{wB} = p^{wC}$ arises, whereas when A discriminates with its tariff policy ($\tau^B \neq \tau^C$) - like in the case of a regional integration agreements - there are different world prices: $p^{wB} \neq p^{wC}$. (1) like (2) is the condition for an interior solution - that is, A trades simultaneously with both its trading partners. Otherwise A would be just trading with the source-country whose good can be offered at a cheaper local price, as it is the only good for which demand is positive.

Note that when A is a small country, its volume of trade are too limited to affect p_x^{*B} and p_x^{*C} which are taken as given. Assuming C is the lowest cost supplier ($p_x^{*C} < p_x^{*B}$), (1) can not possibly hold with a MFN tariff, leading only to bilateral trade between the home country and the cheapest source country.

When A is a large country, p_x^{*B} and p_x^{*C} are endogenously determined and depend on the volume exchanged. In this context, C being the lowest cost supplier just means that it has a grater comparative advantage relative to B and can therefore supply more quantities at the same price. A allocates optimally its imports between the two foreign countries so that the premium per good x (the excess of the tariff-distorted domestic price over the world price) is the same across locations. If this is not the case (e.g. $p_x - p_x^{*B} < p_x - p_x^{*C}$), the tariff revenue per unit of good x will be greater from importing additional units of good x from C and fewer from B . Such reallocation will continue until either the premia is driven to equality or

until all imports come necessarily from a single source.

An MFN tariff further implies that import mix from the two possible sources have to equalize producer prices.

Production

Let $Q_k = Q_k(p)$ and $Q_k^{*i} = Q_k^{*i}(p^{*i})$, $k = x, y$ and $i = B, C$ denote the optimal production quantities of good k respectively in the home country and in foreign country i .¹² The optimal production quantity in each country is determined by the tangency point on the production possibility frontier between the marginal rate of transformation between x and y and the local relative price.

Consumption, Tariff Revenue and GDP

Consumption in each country is a function of the local relative price and income. Given that the tariff/quota revenue is redistributed back to the consumers, national income for the home country (foreign country i) will be the sum of the value of domestic production (GDP) and tariff revenue R (R^{*i}) expressed in units of the local export good at local prices:¹³

$$I(p, R) = pQ_x(p) + Q_y(p) + R \quad (3)$$

¹²Think of the production quantities as the result of profit maximization of the representative firm in each sector k . With the assumption of perfect competition, the amounts produced in each industry will maximize the value of GDP at the local prices. That is, $Q_k(p) = \arg \max_{Q_k} \{pQ_x(p) + Q_y(p) \text{ s.t. } PPF\}$ where PPF is the production possibility frontier. Analogously, $Q_k^{*i}(p) = \arg \max_{Q_k} \{Q_x(p^{*i}) + Q_y(p^{*i})/p^{*i} \text{ s.t. } PPF\}$

¹³Think of the consumed quantities in each country as derived from the maximizing of the representative consumer's preferences subject to the budget constraint. For instance, the representative consumer of foreign country i solves:

$$\begin{aligned} & \max_{C_x^{*i} \geq 0, C_y^{*i} \geq 0} U(C_x^{*i}, C_y^{*i}) \\ & \text{s.t. } p_x C_x^{*i} + p_y C_y^{*i} = p_x Q_x^{*i} + p_y Q_y^{*i} + R^{*i} \end{aligned}$$

$$I^{*i}(p^{*i}, R^{*i}) = Q_x(p^{*i}) + Q_y(p^{*i})/p^{*i} + R^{*i} \quad (4)$$

Given that each country's income is a function of local relative prices and revenue from protection, consumption of good k can be represented by $C_k^{*i} = C_k^{*i}(p^{*i}, R^{*i})$ for foreign country i and by $C_k = C_k(p, R)$ for the domestic economy for $k = x, y$. The tariff revenue in foreign country i in units of the local export good at local prices is defined implicitly by:

$$\begin{aligned} R^{*i} &= (t^{*i}/p^{wi})[C_y^{*i}(p^{*i}, R^{*i}) - Q_y^{*i}(p^{*i})] \\ &= [1/p^{*i} - 1/p^{wi}][C_y^{*i}(p^{*i}, R^{*i}) - Q_y^{*i}(p^{*i})] \\ &\equiv R^{*i}(p^{*i}, p^{wi}) \end{aligned}$$

Denoting by $M^{*i}(p^{*i}, p^{wi}) \equiv C_y^{*i}(p^{*i}, R^{*i}(p^{*i}, p^{wi})) - Q_y^{*i}(p^{*i})$ the imports of good y for foreign country i and, similarly by $E_x^{*i}(p^{*i}, p^{wi}) \equiv Q_x^{*i}(p^{*i}) - C_x^{*i}(p^{*i}, R^{*i}(p^{*i}, p^{wi}))$ foreign country i 's exports of good x , the expression for tariff revenue simplifies to:

$$R^{*i} = (t^{*i}/p^{wi})M^{*i}(p^{*i}, p^{wi}) = [1/p^{*i} - 1/p^{wi}]M^{*i}(p^{*i}, p^{wi}) \quad (5)$$

Consumption in the domestic country is affected by R which originates from trading with both trading partners. Moreover, if A imposes a discriminatory trade policy, its tariff revenue will depend upon the total volume of x that it imports and the composition of this volume across the foreign trading partners. Defining bilateral trade shares by:

$$s^{*i} = \frac{E^{*i}(p^{*i}, p^{wi})}{\sum_{j \in \{B, C\}} E^{*j}(p^{*j}, p^{wj})} \equiv s^{*i}(p^{*B}, p^{*C}, p^{wB}, p^{wC}) \quad (6)$$

and country A 's multilateral terms of trade as:

$$T = \sum_{i \in \{B, C\}} s^{*i} p^{wi} \equiv T(p^{*B}, p^{*C}, p^{wB}, p^{wC}) \quad (7)$$

the revenue from trade protection in units of the local export good at local prices can be defined implicitly by:

$$\begin{aligned} R &= \sum_{i \in \{B, C\}} [C_x(p, R) - Q_x(p)] s^{*i} p^{wi} \\ &= \sum_{i \in \{B, C\}} [C_x(p, R) - Q_x(p)] s^{*i} (p - p^{wi}) \\ &= [C_x(p, R) - Q_x(p)] (p - \sum_{i \in \{B, C\}} s^{*i} p^{wi}) = [C_x(p, R) - Q_x(p)] (p - T) \\ &\equiv R(p, T) \end{aligned}$$

s^{*i} is country A 's share of imports from foreign country i and T is an import share-weighted average of the bilateral terms of trade. A reduction in T represents an improvement in A 's multilateral terms of trade as much as a reduction in p^{wi} represents an improvement in its bilateral terms of trade with foreign country i . (2) implies $T = p^w$ under MFN tariffs. Letting $M(p, T) \equiv C_x(p, R(p, T)) - Q_x(p)$ and $E(p, T) \equiv Q_y(p) - C_y(p, R(p, T))$ be respectively A 's total imports (of good x) and total exports (of good y), the budget constraint of each country implies that trade is balanced, so that:

$$M^{*i}(p^{*i}, p^{wi}) = p^{wi} E^{*i}(p^{*i}, p^{wi}) \quad i = B, C \quad (8)$$

$$TM(p, T) = E(p, T) \quad (9)$$

Note that for each foreign country, the production, consumption, import and export quantities depend exclusively on the tariffs and the world prices. Indeed, once the latter are determined, local prices can be determined too. Analogously, each quantity at home can be expressed ultimately as a function of the local relative price p and the multilateral terms of trade T .

Equilibrium with tariffs

Given a set of tariff $\boldsymbol{\tau} = (\tau^B, \tau^C, \tau^{*B}, \tau^{*C})$ the equilibrium world prices p^{wB} and p^{wC} can be pinned down by (2) together with the following condition:

$$M(p, T) = \sum_{i \in \{B, C\}} E^{*i}(p^{*i}, p^{wi}) \quad (10)$$

(10) is the market clearing condition for good x stating that the world demand for good x is equal to its world supply and recall (2) is the condition for an interior solution. By Walras law, (10), (8) and (9) ensure the market for good y clears as well¹⁴. (10) and (2) determine the two unknown world prices which together with the tariffs determine the equilibrium values of all local prices. In turn, all other quantities (domestic consumption and production, import and export volumes, tariff revenue) and the multilateral terms of trade can be derived.

Given the general form of the functions assumed, it is possible that anomalous but not

¹⁴The market clearing condition for market y is $E(p, T) = \sum_{i \in \{B, C\}} M^{*i}(p^{*i}, p^{wi})$.

impossible cases - such as the Metzler and the Lerner paradox - arise. The Metzler paradox comprises a situation in which the domestic price of an imported good fall after a country levies a tariff on that good. This means that the sector which receives a greater protection is worse off after the introduction of the tariff, so that resources move away from the protected sector since it becomes less profitable. The Lerner paradox refers to the situation in which the terms of trade moves against the tariff-imposing country. In other words, the shortage of demand (excess of supply) that a country induces by imposing a tariff on the imported good does not yield a fall in the international price of that good and consequently an improvement of the country's terms of trade. Technically, this cases can arise because the offer curve is not necessarily well behaved (namely, monotonically increasing and concave to its import axis) even if we assume underlying well behaved supply and demand schedules.¹⁵ I shall restrict my focus to an equilibrium which lies on the upward-sloping segment of the offer curve - the so called "elastic" region.¹⁶

Therefore, I shall assume $i) \partial p^{wi} / \partial \tau^i \leq 0 \leq \partial p^{wi} / \partial \tau^{*i}$ which ensures the terms of trade are always an increasing function of each country's own tariffs (no Lerner paradox). In

¹⁵For instance, an offer curve for foreign country i can be derived solving:

$$\max_{M^{*i}, E^{*i}} B(M^{*i}, E^{*i}) \quad s.t. \quad (8)$$

where

$$B(M^{*i}, E^{*i}) \equiv \max_{C_x^{*i}, C_y^{*i}, Q_x^{*i}, Q_y^{*i}} [U(C_x^{*i}, C_y^{*i}) : M^{*i} = C_y^{*i} - Q_y^{*i}, E^{*i} = Q_x^{*i} - C_x^{*i}, PPF]$$

and PPF is the production possibility frontier.

¹⁶The "elasticity" terminology comes from the fact that the offer curve has a positive slope if and only if the home country's elasticity of demand for imports (with respect to the relative price of the importable) exceeds unity. When the price of a good increases (say because a tariff on this good is levied), it produces three effects: *i*) a production effect (resources move in the sector whose good price has increased) *ii*) a substitution effect (consumer switch demand toward less expensive goods) *iii*) an income effect (due to the improved real income of a country associated to higher terms of trade). Effects *i*) and *ii*) both contribute to decrease imports of this good and increase exports of other goods. On the contrary, effect *iii*) contributes to raise imports of all goods. Restricting to the upward sloping segment of the offer curve means focusing on a situation where the income effect *iii*) is dominated by *i*) and *ii*), so that an increase of the price of the imported good leads to a reduction of its imported quantity.

RTAs formation and trade policy

case of MFN tariffs ($\tau^C = \tau^B = \tau$), these restrictions are simply $\partial p^w / \partial \tau \leq 0 \leq \partial p^w / \partial \tau^{*i}$. Moreover, I shall assume that when other countries raise tariffs on one another, foreign country i 's terms of trade improve - i.e. *ii*) $\partial p^{wi} / \partial \tau^j \geq 0$ ($i \neq j$) and *iii*) $\partial p^{wi} / \partial \tau^{*j} \geq 0$ ($i \neq j$). Finally *iv*) $dp / d\tau^i > 0$ and *v*) $dp^{*i} / d\tau^{*i} \leq 0$, rule out the Metzler paradox and imply a restrictive trade policy always raise local prices. Recall that $p = \tau^i p^{wi}$, so that the total differential of this expression yields the change in the domestic price induced by a tariff change, namely $dp / d\tau^i = p^{wi} + \tau^i dp^{wi} / d\tau^i$. Given that the last term is negative (since an increase in τ^i improves A 's terms of trade), ruling out the Metzler paradox (i.e. $dp / d\tau^i \geq 0$) is equivalent to assume that the terms of trade effect induced by the tariff are not so strong to offset the increment that a tariff has on the domestic price of the good.

Note that the balanced trade conditions together with the market clearing conditions allow to rewrite the domestic tariff revenue as:

$$\begin{aligned} R &= \sum_{i \in \{B, C\}} t^i p^{wi} E^{*i}(p^{*i}, p^{wi}) = \sum_{i \in \{B, C\}} (p - p^{wi}) E^{*i}(p^{*i}, p^{wi}) \\ &= (p - T)M(p, T) \end{aligned} \quad (11)$$

where the intuition for the first equality follows from the bilateral nature of the trading relations which make foreign country i 's exports necessarily the home's imports from i . Note R^{*i} and R are an increasing function of foreign country i 's terms of trade, under our assumption that goods are normal.

Furthermore, given that tariff revenues are only functions of local relative prices and world relative prices, also national incomes (3) and (4) are only functions of local and world

relative prices.

Equilibrium with volume quotas

The bilateral nature of exchanges between A and each foreign country i fixes B 's and C 's exports of good x at the quantity level set by A 's import quota of m^B units of good x from B and of m^C units from C . Thus, the equilibrium world prices are determined by

$$E^{*i}(p^{*i}, p^{wi}) = \bar{m}^i \quad i = B, C. \quad (12)$$

(10) becomes a tautology because A 's total imports are necessarily the sum of the two binding volume quota restrictions. In addition, if foreign country i imposes a volume quota of \bar{m}^{*i} on its imports, world price p^{wi} can be explicitly determined by (8) as the ratio \bar{m}^{*i}/\bar{m}^i .

When the only restrictions in place are A 's import quotas on imports from country C , $E^{*C}(p^{*C}, p^{wC}) = \bar{m}^C$ still determines p^{wC} and (10) determines the remaining unknown p^{wB} .

This highlights a specificity of a quota with respect to a tariff which is worth mentioning as it will play a crucial role in the welfare analysis. If a quantitative restriction on imports from C remains binding even after a reform that abolishes a formerly existing quota m^B imposed on imports from B , $E^{*C}(p^{*C}, p^{wC}) = \bar{m}^C$ will still fix the level of imports from C and the terms of trade at the pre-reform level. Thus, all the adjustments induced by the reform affect only the bilateral relations among B and A . On the contrary, a reform that alters a pre-existing A 's MFN tariffs lowering only the tariff risen on imports from B will lead to world price adjustments (see (2) and *i*) and *ii*) above).¹⁷ Intuitively, if the home

¹⁷An example would be a reform that lowers τ^B to 1 and leaves τ^C unaltered at τ .

government taxes more heavily the exports of foreign country C , then the home demand for exports from C is reduced whereas the home demand for exports from B is increased, resulting in a terms of trade loss for C and a terms of trade gain for B . In this sense, the tariff protection is more likely to induce cross market effects than quota protection.

This insight gives the intuition for a potentially more severe trade diversion in tariff protected than in quota restricted markets. Note, however, it is a feature of country-specific volume quota. Were A restricting only its total imports of good x to $\overline{M} = \overline{m}^B + \overline{m}^C$ units (global quota), the outcome would be similar to the tariff case. Imports from each source country would be allocated such that the quota premium is equalized across source country ($p_x - p_x^{*B} = p_x - p_x^{*C}$) or all imports come from the single source that maximizes the quota premia. Indeed, there are several combinations of import-mixes from the two source countries that can satisfy (10) with M held fixed at \overline{M} and the one chosen is maximizing the quota premium. Thus, country-specific direct quantitative restrictions are more restrictive than an equivalent global quota because they restrict the import mix to a specific one which is not necessarily the optimal one.¹⁸

3 Regional Integration with different trade policies

I shall consider now the welfare outcome of a regional integration (RI) policy under alternative trade policies. Without loss of generality, country A has the option to sign a trade

¹⁸Country-specific quotas are widely employed. The United States, for instance, imposes limitations on imports of sugar on a country-by-country basis, with each specific exporting country having its own quota allotment of sugar exports. Trade in textiles and apparel is managed in much the same way; importing nations negotiate country-specific import targets with each significant exporter. See Markusen et. al (1995).

agreement with B reducing preferentially trade barriers to the member country (B), but leaving unaltered those raised vis-à-vis non-member countries (C).

The desirability of a RIA has to be evaluated according to some welfare measure and I shall assume national governments maximize national income. This choice corresponds to national governments maximizing the indirect utility function v of the representative consumer. Therefore, $V(p, T) = v(p, I(p, T))$ measures A 's welfare and, similarly $V^{*i}(p^{*i}, p^{wi}) = v(p^{*i}, I^{*i}(p^{*i}, p^{wi}))$ represents the welfare of foreign country i .¹⁹

Countries may trade in two different policy regimes. In the first one, all countries are making use of tariff measures to protect national industries - as in the 70s. The initial tariff vector is $\tau_I = (\tau_I^B, \tau_I^C, \tau_I^{*B}, \tau_I^{*C})$, I for *initial*. Then, A and B join into a FTA which gives free access to the internal market only to member countries, leaving trade barriers against non member countries unaltered (reform t hereafter). Thus, the post reform tariff vector (indexed by t) is $\tau_t = (1, \tau_I^C, 1, \tau_I^{*C})$. That is, A and B trade freely one another. Given that C is not part of the agreement and its tariff level is unchanged after the reform, for simplicity I shall set $\tau^{*C} = 1$, meaning it is trading freely before and after the reform. Therefore, $\tau_I = (\tau_I^B, \tau_I^C, \tau_I^{*B}, 1)$ and $\tau_t = (1, \tau_I^C, 1, 1)$.

Consequently, let me denote as V_I the welfare associated to the pre-agreement equilibrium and likewise, V_t the welfare associated to the post-agreement equilibrium when trade is restricted by tariffs.

¹⁹Bagwell and Staiger (2002) show that government preferences can be generalized to allow for a wide range of economic and political motivations.

In the second one - as in the 90s - trade is initially restricted by direct quantitative restrictions and countries A and B undergo through the same kind of piecemeal reform which liberalizes the internal market to member countries while it leaves unchanged the quantity restrictions vis-à-vis non-member countries (reform q hereafter). Therefore $\bar{m}_s = (\bar{m}_s^B, \bar{m}_s^C, \bar{m}_s^{*B}, \bar{m}_s^{*C})$ (s for *start*) is the pre-agreement quota vector and $\bar{m}_q = (no, \bar{m}_s^C, no, \bar{m}_s^{*C})$ is the post-reform quota vector. *no* in the first element of this vector means that no volume quota is in place in country A for imports of good x from B . As above, I shall assume for simplicity that C is in a free trade regime. We have, $\bar{m}_s = (\bar{m}_s^B, \bar{m}_s^C, \bar{m}_s^{*B}, no)$ and $\bar{m}_q = (no, \bar{m}_s^C, no, no)$. Thus, the volume quota faced by C remains the only restriction in place after the FTA is signed. Obviously, V_s denotes the welfare associated to the pre-agreement equilibrium and likewise, V_q the welfare associated to the post-agreement equilibrium when trade is restricted by volume quota.

As well established in the literature, such partial reforms should be evaluated in terms of terms of trade and efficiency effects. As shown in Grinols and Wong (1991) and Ju and Krishna (2000a), reform q can ensure Pareto gains (s0, $V_q \geq V_s$) through a system of self-financing lump sum transfers provided that the following condition holds:²⁰

²⁰This formula is derived in the appendix. See also Feenstra(2004). The intuition is that the revenue R collected after the implementation of the reform can be used to finance a system of lump sum transfers which compensates losers and taxes gainers. This redistributive policy aims at equilibrating the uneven benefit distribution arising with the trade reform and therefore can achieve Pareto gains for all consumers.

$$\begin{aligned}
 & (p_s^{wB} - p_q^{wB})E^{*B}(p_s^{*B}, p_s^{wB}) + (p_s^{wC} - p_q^{wC})E^{*C}(p_s^{*C}, p_s^{wC}) \\
 & + (p_q - p_q^{wB})[E^{*B}(p_q^{*B}, p_q^{wB}) - E^{*B}(p_s^{*B}, p_s^{wB})] \\
 & + (p_q - p_q^{wC})[E^{*C}(p_q^{*C}, p_q^{wC}) - E^{*C}(p_s^{*C}, p_s^{wC})] \geq 0
 \end{aligned} \tag{13}$$

that is, regionalism integration must avoid adverse terms of trade or efficiency effects to ensure Pareto gains. The first line in (13) indicates the impact of the reform on the terms of trade: if the RTA reduces (increases) the price of imports (exports), it will contribute positively to raise welfare. The last two lines are the change in tariff revenue evaluated at the post-reform tariff vector (see (11)) and represent efficiency gains.²¹ Trade diversion means that A switches the supplier from outside the FTA (where tariff revenue is collected) to another within the FTA (with no tariff revenue), so that tariff revenue falls. The last two lines would then be negative. Thus, RTAs should reduce trade diversion as much as possible to be welfare enhancing. It is worth mentioning that (13) is a sufficient condition, so that its violation does not mean the reform under evaluation is a failure a priori, but it just needs a careful examination.

It is instructive to consider first the small open economy, first. It allows to abstract from terms of trade and focus closely only on the different degree of trade diversion implied by the two different instruments of trade policy. Then, I shall analyze the large country case

²¹See Feenstra (2004). The tariff reflects the difference between domestic and international prices: if positive, that indicates that marginal costs at home exceed international prices, so it would be more efficient for the country to import the good. Therefore the term $(p_q - p_q^{wi})[E^{*i}(p_q^{*i}, p_q^{wi}) - E^{*i}(p_s^{*i}, p_s^{wi})]$ gives a measure of the efficiency gain by attracting imports towards the highest-tariff sectors after the implementation of a reform.

where the same mechanisms of trade diversion interact with terms of trade effects, making the analysis more cumbersome.

Given that the literature on regional integration has found trade diversion theoretically and empirically relevant, the interesting case is represented by a potential trade diverting agreement.

Therefore, I shall consider C - the outside union member - as the lowest cost supplier or the cheapest source country in the small open economy. In the large economy, C has a comparative advantage in good x relative to B , translating in greater export supply for any given world price. Note that trade diverting agreements do not need to be necessarily welfare decremental, but they can be welfare enhancing.²² In particular, I am interested in whether such agreements are welfare enhancing when trade protection is achieved through NTBs, so that the formation of RIAs in the 90s can be related to the observed change in the nature of trade protection.

3.1 The small open economy case

The small open economy is illustrative of how a volume quota has different effects from tariff when a country undergoes a piecemeal tariff reforms.

The assumption of a small open economy ensures that:

- A 's trade policy does not affect the terms of trade with the two possible suppliers, B and C , but only the relative domestic prices $p \equiv p_x/p_y$. As world prices do not change

²²Trade diverting preferential agreement are necessarily welfare reducing under two specific assumptions: i) no substitution in consumption and ii) a linear production possibility frontier. See Markusen et. al. (1995).

in response of a trade reform, I shall denote them as p^{wB} and p^{wC} without any subindex (I, s, t, q) .

- B 's and C 's supply is infinitively elastic at the world relative price p^{wi} , $i = B, C$ for trade with A , with $p^{wC} < p^{wB}$, as C is assumed to be the lowest cost supplier,. Assume further that $p^{wC} \leq p^{wB} \leq \tau_I^C p^{wC} = p$, that is that B 's terms of trade with A is lower than the tariff-distorted domestic price of good x coming from C .²³

To see how the two types of instruments lead to different adjustment mechanisms when a RIA is joined, suppose country A is initially protecting its industries by a MFN ad valorem tariff $t_I = t_I^B = t_I^C$. Imports from B would be more expensive than imports from C (by assumption $p^{wB} \geq p^{wC}$), and could be sold domestically only at a higher price ($\tau_I p^{wB} \geq \tau_I p^{wC}$) - so that (2) is violated and A will be only trading with C . Indeed, given that C is willing to supply all $M(p_I, p_I^{wC})$ at the world price p^{wC} , all imports of x come from C .²⁴ Thus, in equilibrium there are only bilateral relations among A and C , but the tariff revenue is positive and given by (11).²⁵

However, when A grants preferential access to its market only to country B and retains the same tariff on imports from C , the relative domestic price of good x imported from B becomes $p_t = p^{wB}$, lower than the price $\tau_I p^{wC}$ at which each unit imported from C could possibly be sold in A . Given that A 's total import demand at p^{wB} is completely satisfied,

²³This assumption puts A into the situation of a trade diverting agreement.

²⁴In other words, (10) is $M(p_I, p_I^{wC}) = E^{*C}(p_I^{*C}, p_I^{wC})$ and $E^{*B}(p_I^{*B}, p_I^{wB}) = 0$. As shown in section (2), (2) can not possible hold in the small open economy with MFN tariff. The equilibrium is necessarily a corner solution where A trades only with partner C .

²⁵Obviously $E^{*B}(p_I, p_I^{wB}) = 0$ in (??), as the equilibrium is in the corner solution where A trades only with C .

trade is completely diverted from country C to country B . In this case, in equilibrium only A and B trade each other and the tariff revenue is lost because all imports come from a union-members to which no tariff is applied.

The outcome differs under the NTBs regime. Suppose that A initially sets a global quota \bar{M}_s on the total number of units of x imported. Assume that it is binding and, for expositional simplicity, it is the tariff-equivalent quota (i.e. $\bar{M}_s = E^{*C}(p_I^{wC})$), that is the implicit ad valorem tariff associated to it is t_I .²⁶ Therefore, the initial equilibrium is formally identical to the the initial tariff-ridden equilibrium with all units imported exclusively from C and sold domestically at the tariff-distorted domestic price $p_s = p_I = \tau_I p^{wC}$. Therefore, the tariff revenue collected is the same too.

After A joins into a RIA with B , A will be trading freely with B , but it can also keep trading with C within the limits imposed by the volume quota \bar{M}_s , now imposed exclusively on imports from C . Note that C faces an import restriction as high as it was before the agreement, while B benefits from the removal of all trade barriers. Since the international prices at which country A can make international transactions are not affected by this policy, a quota-license holder can make positive profit by buying x from the lowest cost supplier C at p^{wC} and re-sell it at the higher domestic price $p_q = p^{wB}$. All license holders will have an incentive to use entirely their license and the amount of x imported from C is therefore \bar{M}_s .

This is an equilibrium because C is willing to supply all the \bar{M}_s units at p^{wC} , whereas B

²⁶The equivalence result between a tariff and a quota due to Bhagwati (1965) holds - in the sense that a tariff rate will produce an import level which, if alternatively set as a quota, will produce an identical discrepancy between foreign and domestic prices.

serves the residual import demand $(M_q - \overline{M}_s)$ at p^{wB} .²⁷ All three countries end up trading simultaneously and therefore some revenue $R_q = (p^{wB} - p^{wC})\overline{M}_s$ from industry protection can be maintained even after the implementation of the agreement. The part of the original revenue that is given up $(p_I - p^{wB})\overline{M}_s$ reflects exactly the reduced economic value of the quota which is not as stringent as it was before the integration.²⁸ Indeed, those units of imports desired by country A in excess of the quota limit that could not be imported before the agreement, can now be imported from Country B .

Although the partial reform implemented under the two different types of protection has identical effects on A 's relative domestic prices (i.e. $p_q = p_t = p^{wB}$), the composition of the volume of trade across the two supplier is rather different as trade diversion does not occur under quantitative restrictions to trade. With quota protection, A expands its volume of trade with B (trade creation), but trades with C ex-post as much as it did ex-ante the agreement, whereas with tariff protection, trade diversion alters the pre-union trade patterns between A and C . Only the FTA implemented in presence of quota restrictions preserves the imports and the exports with the rest of the world and results therefore welfare improving as shown by Krishna and Panagariya (2002). This observation is the essence of the different welfare consequence of a trade agreement under the two types of trade policy instruments.

This different degree of trade diversion implied by the two types of trade barriers will extend naturally to the large country case. Therefore I shall defer the formal proof of this result till later in the more general case of large countries.

²⁷The binding quota ensures the desired imports of good x necessarily exceed the \overline{M}_s units after the agreement is in place.

²⁸The pre-agreement quota rent is $R_I = (p_I - p^{wC})\overline{M}_s = [(p_I - p^{wB}) - (p^{wB} - p^{wC})]\overline{M}_s$.

Note that such conclusion would still hold if A were initially setting country specific import restrictions \bar{m}^B and \bar{m}^C . After a RTA that abolishes just \bar{m}^B , but it leaves unchanged the quota raised on imports from C at \bar{m}^C , A 's post-agreement purchases from C are unchanged provided that this quota is still binding.

3.2 The large economy case

Although a small country always gain from reform q , a large country may not do so if such reform leads to adverse terms of trade effects. Let p_s^{wB} and p_s^{wC} the world price in the pre-agreement equilibrium determined by (12). Note that E^{*C} is an increasing function of only p^{wC} as C does not protect its imports (i.e. $p_s^{*C} = p_s^{wC}$). Supposing \bar{m}_s^C keeps be binding after the reform, (12) implies that the terms of trade with C are unchanged (i.e. $p_q^{wC} = p_s^{wC}$) and, in turn, $M^{*C}(p_q^{wC}) = M^{*C}(p_s^{wC})$ by (8). (10) implies that also the terms of trade with B are fixed at their initial value (i.e. $p_q^{wB} = p_s^{wB}$), as proved in the appendix.²⁹ Therefore, after the reform is implemented, A is trading with partner C as much as it was doing before the reform was approved, importing and exporting the same quantities. This is the criteria for welfare enhancing Free Trade Areas proposed by Krishna and Panagariya (2002). The preservation of trade volumes with non-members countries guarantees that their terms of trade and, in turn, their welfare is preserved (i.e. $V_s^{*C} = V_q^{*C}$) too. To see that it also improves the welfare of participating government, it is easy to verify that (13) is satisfied

²⁹To see this point, note that (10) implies that $M(p_q, T_q) - E^{*B}(p_q^{*B}, p_q^{wB}) = E^{*C}(p_q^{wC}) = \bar{m}_s^C = M(p_s, T_s) - E^{*B}(p_s^{*B}, p_s^{wB})$. Given market x is intially clearing at (p_s^{wC}, p_s^{wB}) , it necessarily keeps clearing at these world prices after reform q . That is, $p_s^{wC} = p_q^{wC}$, $p_s^{wB} = p_q^{wB}$. Indeed, the increase in B 's export supply matches exactly with A 's greater total import demand (i.e. $M(p_q, T_q) - M(p_s, T_s) = E^{*B}(p_q^{*B}, p_q^{wB}) - E^{*B}(p_s^{*B}, p_s^{wB})$), so that market x necessarily clears at the same pre-reform world prices.

being equal to zero. The first line is 0 because the world price vector is unchanged and, therefore countries in the FTA do not experience terms of trade gains or loss. The second line is 0 as no quota premium arises from FTA members ($p_q^{*B} - p_q^{wB} = 0$). Finally, the last line is 0, as trade diversion does not occur and import volumes with third countries are preserved (i.e. $E^{*C}(p_q^{*C}, p_q^{wC}) = \overline{m}_s^C = E^{*C}(p_s^{*C}, p_s^{wC})$). Clearly, this agreement increases B 's volume of trade, contributing to raise its welfare too - see the appendix (7.1).³⁰ Henceforth, the agreement is mutually favorable. We have the following proposition.

Proposition 1 *A RIA between the home country and foreign country i that abolishes volume quota one another leaving unaltered former volume import restrictions vis-à-vis non-member country j is welfare enhancing, provided that such import restrictions keep being binding after the agreement is implemented.*

A FTA implemented in presence of quota restricted trade does not yield terms of trade gains, but it also avoids any trade diversion. This is enough to ensure the possibility to achieve Pareto gains through a system of transfers among consumers and, more importantly, among countries. The intuition for this result is suggested by the literature on piecemeal tariff reform.³¹ Since following reform q , some binding quota are relaxed with others remained fixed and there are no other distorted markets, the welfare change is positive because the relaxation in any quota unambiguously enlarges the economy's consumption set. Importantly, these gains from consumption are not offset by adverse terms of trade effects.

³⁰It can be directly verified that (24) in the appendix - a formula like (13) for foreign country $*B$ - holds. The first term is 0 since reform q has no terms of trade effect. The second term is 0 as no quota is risen on imports from member countries (i.e. the implicit tariff is $t^{*B} = 0$). If B had multiple partners too, trade patterns would be symmetric to A , exporting x to both the union country A and a third country outside the FTA and importing y from a union country A and the third country outside the union. Thus, the question of whether union formation is welfare improving for each member country could be analyzed in the same way for both members of the union. Since the FTA results welfare-enhancing for A , it would be so for B too.

³¹See Vousden (1990), chapter 9 pag pag 217.

I think of this proposition as an indication of the desirability of regionalism in presence of quota-restricted trade. This is known as the "*compensation principle*" due to Chipman (1987) - in the sense that if regional integration agreements combined with lump-sum transfers could make everyone better off, then we accept such reform itself as a worthwhile policy, even when the transfers are not made. Indeed, the knowledge required to implement the long list of lump sum compensatory payments, some of them international in nature, make the practical implementation of these transfers very difficult.

Obviously, proposition (1) holds for a small country too, as (13) is satisfied by the small country above.

On the contrary, reform t is not necessarily welfare improving because the trade diversion induces terms of trade movements as well. The initial equilibrium world prices p_I^{wC} , p_I^{wB} solve (10) and (2). After the reform t , the elimination of tariff within the FTA reduces the local price of the imported good in the two member countries, so that $[1/\tau_I^{*B} \leq p_t^{wB}/p_I^{wB} \leq \tau_I^B]$, whereas by (2), C 's terms of trade worsen ($p_t^{wC} \leq p_I^{wC}$) to preserve competitiveness in the union market.³² This appreciation of A 's terms of trade against C originates from the trade diversion of imports formerly coming from the outside member country in favor of intra-union imports (i.e. $E^{*C}(p_t^{wC}) \leq E^{*C}(p_I^{wC})$). The elimination of tariff on union trade shifts out both A 's import demand and B 's export supply. Depending on this relative shifts, B 's terms of trade may appreciate or depreciates. First, suppose it appreciates ($1 \leq p_t^{wB}/p_I^{wB} \leq \tau_I^B$) - for instance because at the initial world price there is an excess demand of good x by A to be

³² $[1/\tau_I^{*B} \leq p_t^{wB}/p_I^{wB} \leq \tau_I^B]$ follows from $p_t = p_t^{wB} \leq \tau_I^B p_I^{wB} = p_I$ and $p_t^{wB} = p_t^{*B} \geq p_I^{*B} = p_I^{wB}/\tau_I^{*B}$. $p_t^{wC} \leq p_I^{wC}$ follows from $\tau_I^C p_t^{wC} = p_t \leq p_I = \tau_I^C p_I^{wC}$. This is fully consistent with price restriction *i*) and *ii*).

cleared out. Replacing q with t and s with I , (13) is not necessarily positive: the second line is 0 as no tariff is levied on member's imports, whereas the third line is negative as trade diversion reduces tariff revenue from non-members countries. Finally, the first term in the first line is negative while the second term is positive. For welfare to raise, favorable terms of trade effects against non-member countries have to be strong enough to compensate for both terms of trade loss vis-à-vis member countries and efficiency loss.

Supposing instead that $(p_t^{wB} \leq p_I^{wB})$, welfare gains are more likely as A experiences favorable terms of trade effects with both member and non-member trading partners which can offset the efficiency loss associated to trade diversion (the negative term in the last line of (13)).

Note for the small open economy above, Pareto gains can not be ensured either, since (13) is violated. The first two lines are clearly 0, whereas the last line is negative because of trade diversion. The intuition is again that trade diversion associated to the preferential reduction of tariffs may not necessarily translates into the enlargement of the economy's consumption set, especially when imports are just diverted from one source to another. Moreover, adverse terms of trade plausibly associated to trade diversion make imports more expensive and reduce welfare.

It is worth pointing out a number of interesting features of this result.

First, the desirability of reform t has to be evaluated empirically case by case, as widely recognized in the literature. For instance, Grinols (1984) finds that Great Britain incurred in a welfare loss when it joined the European Economic Community in 1973 and (13) was

evaluated to be negative. In contrast, proposition (1) suggests a general desirability of reform q which can provide an explanation for a renewed interest in RIAs policy in the 90s. Indeed, policy makers are usually working in a second best world, attempting to reduce some distortions while others remaining firmly in place. As we know from the theory of second best, such changes do not necessarily increase welfare. However, it would be helpful to have some simple rules - a rule of thumb - telling them "which way is up". Then, Proposition (1) suggests reform q can be readily understood and implemented by policy makers with some confidence that a welfare improvement will be the outcome.³³

Second, because "the way up" is so clear when all trade barriers consists of quantitative restrictions, it would appear that the difficulty of identifying welfare-enhancing FTA can be resolved simply by converting all explicit taxes or subsidies to quotas and then gradually relaxing the quotas. Such an approach runs counter to the view of reform which is popular to most policy makers and adopted by the WTO in 1995, that is change quotas to tariffs and then set about reducing the tariffs - so called "tariffication" process. However, quotas have many other disadvantages relative to tariffs, particularly in the presence of monopoly and/or foreign retaliation, and such considerations presumably are what motivate the popular view.³⁴

Third, the welfare outcome of reform q and reform t is different, though both reforms leave unaltered former trade barriers against non-member countries. This is the criteria established

³³This has some analogies with the large literature on piecemeal tariff reforms. A number of economists have looked for simple piecemeal reform rules that could be leading to welfare enhancing outcomes with some confidence. Examples of such rule of thumb rules are proportional reduction of all trade taxes or reduction of the highest tariff rate first. See Vousden (1990), Chapter 9 for a review.

³⁴See Vousden (1990), chapter 4, 5 and 6.

by GATT-WTO regulations to permit preferential trade agreements. Article XXIV of GATT explicitly imposes “the duties (with outside parties) shall not on the whole be higher or more restrictive than the general incidence of the duties...prior to the formation”. This criteria naturally implements the efficient rule for welfare-enhancing FTA in case of NTBs, as it helps to keep purchases from the rest of the world fixed. In contrast, in case of tariff protection (reform t), it is unable to lead to welfare improving FTA as this principle does not suffice, on its own, to preserve trade volumes with non-member countries. As shown by Bagwell and Staiger (2004), the FTA would do so if only complies with two other principles of GATT-WTO - namely reciprocity and MFN. However a FTA could not possibly respect MFN as, by definition, it is a legal recognized exception to it.

Fourth, the model suggests that when a RTA is implemented in presence of NTBs, the terms of trade vis-à-vis non member countries should be preserved. This seems at odd with the empirical findings by Chang and Winters (2002) of favourable terms of trade effects of MERCOSUR for Brazil vis-à-vis US, Japan, Germany, South Korea and Chile, who are all exporting into the union-area. However, Mercosur entailed also tariff reductions which can lead to terms of trade effects as predicted by the model (reform t). Moreover, it is important to note that the preservation of trade volumes with non-member countries is a specificity of country-specific volume quotas. Consequently, terms of trade effects can also occur in presence of only quantitative restrictions, when, for instance, a RTA provides that countries relax a global import volume quota or a formerly employed VERs. Given the empirical relevance of VERs in trade policies, next section will consider them specifically and show

that, indeed, may lead to favourable terms of trade improvements for member countries and loss for non member countries as for Brazil after Mercosur.³⁵

4 RIAs and VERs

A VER is a quota imposed by an exporting country on its exports to another country in response to pressure by the importing country. In my framework, the initial VERs vector $\bar{m}_o = (\bar{m}_s^{*B}, no, \bar{m}_s^B, \bar{m}_s^C)$ where foreign country i imposes a VER of \bar{m}_s^i units of good x , whereas A restricts its exports to \bar{m}_s^{*B} units of good y toward B , mirrors the protection achieved by the initial import-volume quota vector \bar{m}_s . The equivalent of reform q in terms of VERs relaxation (hereafter denoted as reform v) implemented by a bilateral agreement yields the post-reform VER-vector $\bar{m}_v = (no, no, no, \bar{m}_s^C)$. For an exporter in foreign country i , a unit of the exported good x is more valuable on the international market than on the domestic market (i.e. $p_x \geq p_x^{*i}$) due to the artificial scarcity created by a VER. Analogously, for a home exporter, the international price of the exported good y benefits from such artificial restraint and results above its domestic price (i.e. $p_y^{*i} \geq p_y$). Then, the untaxed or "world" relative price at which international transactions are made is given by $\tilde{p}^i = p_x/p_y^{*i}$ which still represents foreign country i 's terms of trade and the ratio of exporter prices. With

³⁵VERs have become most pervasive in the textiles and clothing area, applying to 80% of the world trade in these industries through the various bilateral agreement which constitutes the Multifibre Arrangement (MFA). They are also particularly in evidence in the steel industry, in which they limit steel exports from Japan and the EEC to the USA. They were employed in the automobile industry, in which they restricted exports from Japan to the USA, Germany, France, UK and Italy. For example, after 1981, the Japanese Ministry of International Trade and Industry (MITI) told each auto manufacturer in Japan how much it could export to the U.S. In the late 1980s, US and the European countries rescinded their request to Japan for restricted automobile exports. Later on, they have begun to be applied to a range of electronic consumer goods. As documented by Laird and Yeats (1990), over the period 1981-1986 a major shift has occurred in the use of VERs on import volumes as opposed to other forms of NTBs.

this definition, relative domestic prices are in the same relation with world prices as above (i.e. $p^{*i} \leq \tilde{p}^i \leq p$) and (8) is still foreign country i budget constrain with p^{wi} replaced by \tilde{p}^i .

One major difference with the import quota case analyzed above, is that both E^{*C} and E^{*B} are vertical at \bar{m}_s^C and \bar{m}_s^B when VERs are binding and do not pin down world prices.

These are determined by:

$$M(p_o, \tilde{T}_o) = \bar{m}_s^B + \bar{m}_s^C \quad (14)$$

$$M^{*B}(p_o^{*B}, \tilde{p}_o^B) = \bar{m}_s^{*B} \quad (15)$$

with $\tilde{T} = \sum_{i \in B, C} \tilde{s}^i / \tilde{p}^i$ and $\tilde{s}^i = M^{*i}(p^{*i}, \tilde{p}^i) / M^{*B}(p^{*B}, \tilde{p}^B) + M^{*C}(p^{*C}, \tilde{p}^C)$.³⁶ Note that differently from above, an increase in \tilde{T} represents an improvement in A 's multilateral terms of trade. (15) determines \tilde{p}_o^B while (14) pins down \tilde{T}_o and, in turn, \tilde{p}_o^C .³⁷ Generally, \tilde{p}_o^i will result different from the terms of trade p_s^{wi} because C and B are moving the terms of trade in their favour by imposing a VER whereas, in the import quota case, was A to benefit from the terms of trade appreciation with the imposition of a country-specific quota to each of its trading partners. Moreover, revenue formerly collected with the imposition of such import quotas, accrues under the VER scheme to foreign country i , implying quite different lump sum transfers for consumers of country A . In turn, in general equilibrium models, also relative domestic prices will be generally different in the two situations ($p_s \neq p_o$ and

³⁶To develop an intuition it is useful to think of VERs in terms of implicit export taxes. t^{*i} is the export tax levied by foreign government i on its exports of good x - so that an exporter receives $p_x^{*i} = (1 - t^{*i})p_x$. t^B is the export tax imposed by A on exports directed to B , while $t^C = 0$ - so that an exporter in the home country receives $p_y = (1 - t^B)p_y^{*B}$ when it sells the good in market B and $p_y = p_y^{*C}$ when it sells its good in market C . Export taxes and import tariffs have qualitatively the same effects because they distort domestic prices in the same manner favouring the import-competitive sector.

³⁷Actually \tilde{p}_o^B can be derived more explicitly and it is $\tilde{p}_o^B = \bar{m}_s^{*B} / \bar{m}_s^B$.

$p_s^{*i} \neq p_o^{*i}$) because of income effects associated with changes in the terms of trade.

After the implementation of the RTA between A and B , only the VER imposed by C on its exports is firmly in place and assumed to be binding, which implies that E^{*C} is still vertical at \bar{m}_s^C . Clearly, the export quantities toward A are preserved, but, differently from the import quota case, the terms of trade are not necessarily preserved because, under a binding VER only this quantity is supplied regardless of its price. By (8), the terms of trade would be unchanged if only the quantity imported by C were also unchanged.

The market of good y results free of protection and country A has two potential buyers, namely B and C . Of course, A would maximize its export revenue by selling to the best buyer and will allocate to it all its exports unless the price per unit of good exported is not driven to equality by such allocation (i.e. $p_y^{*B} = p_y^{*C}$). Therefore, the post-agreement equilibrium in which A is trading with both partners is characterized by the following market x clearing condition and arbitrage condition:

$$M(\tilde{T}_v) = E^{*B}(\tilde{p}_v^B) + \bar{m}_s^C \quad (16)$$

$$\tilde{p}_v^B = \tilde{p}_v^C \quad (17)$$

The y market clearing condition $E(\tilde{T}_v) = M^{*B}(\tilde{p}_v) + M^{*C}(p_v^{*C}, \tilde{p}_v)$ implied by (16) and (8) provides us with an intuition for the effects of the RTA. Initially, the allocation of A 's exports to its partners is constrained by the A 's VER of \bar{m}_s^{*B} units of good y (see (15)). Following the trade liberalization, A will wish to export more units and its allocation among

its trading partners is driven by revenue maximization. Very likely, such optimal export mix toward its trading partners is different from the original export mix and leads to terms of trade movements. In particular, it can be shown (see the appendix) that the effect on \tilde{p}_v^B is ambiguous, while \tilde{p}_v^C and M^{*C} are smaller than initially. The preserved exchanges of good x between A and C does not longer suffice to preserve the volumes traded of good y as well as the terms of trade. Therefore, negotiated bilateral agreements for the elimination of VERs also have favourable terms of trade effects for the the negotiating parties vis-à-vis third parties.

As shown in the appendix, it is possible to express the sufficient condition for Pareto gains for the home country as:

$$\sum_{i \in B, C} \left(\frac{1}{\tilde{p}_v^i} - \frac{1}{p_v} \right) [M^{*i}(p_v^{*i}, \tilde{p}_v^i) - M^{*i}(p_o^{*i}, \tilde{p}_o^i)] + \left(\frac{1}{\tilde{p}_v^B} - \frac{1}{\tilde{p}_o^B} \right) M^{*B}(p_o^{*B}, \tilde{p}_o^B) + \left(\frac{1}{\tilde{p}_v^C} - \frac{1}{\tilde{p}_o^C} \right) M^{*C}(p_o^{*C}, \tilde{p}_o^C) \geq 0 \quad (18)$$

where the first line is 0 in my framework because A does not employ any VER after reform v is implemented. Therefore, Pareto gains only depend on terms of trade movements. In particular, when both $\tilde{p}_v^B \leq \tilde{p}_o^B$ and $\tilde{p}_v^C \leq \tilde{p}_o^C$, Pareto gains can be ensured, otherwise possible adverse terms of trade with member countries have to be offset by gains vis-à-vis non member countries. Most remarkably, Pareto gains for country A under a VER reform are more likely than under reform t , although both reforms t and v entail similar terms of trade effects. The reason is that country A had already given up its tariff revenue when it negotiated a VER protection and regionalism (reform v) has no consequences for the loss of

such revenue in this case. In contrast, (13) showed that favourable terms of trade induced by reform t had in any case to compensate for tariff revenue loss (the negative term in the third line) to ensure Pareto gains.³⁸

5 Comparison of the different strategies to RIAs

We have shown that policy makers may be quite confident about pursuing regional integration policies to rise the well being of countries in a context where trade is restricted by NTBs, while the outcome of this strategy is quite case-specific and needs more careful evaluation in a context of tariff restricted trade. However, in many cases, the option for such preferential reforms may still represent an improvement relative to pre-existent tariff-ridden trade relations. In spite of a more generalized desirable welfare outcome of RTAs in presence of NTBs, it would be interesting to be able to compare the size of the gains associated to piecemeal reforms with either type of instruments. In terms of the notation of my model, I am interested in comparing gains from a reform q relative to those arising under reform t (i.e. $V_q - V_s$ vs. $V_t - V_I$). This enables us to construct useful counterfactuals. Would a country joining into a regional trade agreement in the 70s have gained at least as much had it signed the same agreement twenty years after at the end of the 80s? Or would a country recently joining into an agreement in presence of NTBs gained more or less had it enrolled into this agreement much earlier in the 70s when tariff rather than NTBs were in place?

Therefore, these counterfactual are useful to shed some light on how attractive and valid is

³⁸Note that if A were imposing initially a VER to both its trading partner and leaving its VER unaltered against country C after reform v , trade diversion translates into rents loss and gains from regionalism would be less plausible as well.

the option for a RTA in the 90s for a country for which it was not worth joining into an agreement in the 70s.

In such thought experiments, it is important to attribute any change in the welfare only to the piecemeal tariff/quota reform. Clearly, $V_s = V_I$ would ensure A has the same initial welfare with either trade policy and therefore any welfare change can be attributed to the different adjustments implied by volume quota liberalization and tariff liberalization in the process of a RTA reform. The assumption of a competitive market for import quota licenses ensures the equivalence result between a tariff and a volume quota holds.³⁹ Therefore, the levels of prices, production, consumption, imports, exports and transfers of rents to the protected industry prevailing in the pre-reform equilibrium with tariff are identical to those characterizing the pre-agreement equilibrium with NTBs. It follows the welfare associated to both equilibrium must coincide, i.e. $V_s = V_I$. Given this equivalence, $V_q - V_s \geq V_t - V_I$ reduces to study under which conditions A is better off in the q equilibrium as opposed to the t equilibrium ($V_q \geq V_t$) which can still be evaluated similarly to (13) by:

$$\begin{aligned}
 & (p_t^{wB} - p_q^{wB})E^{*B}(p_t^{*B}, p_t^{wB}) + (p_t^{wC} - p_q^{wC})E^{*C}(p_t^{*C}, p_t^{wC}) & (19) \\
 & + (p_q - p_q^{wB})[E^{*B}(p_q^{*B}, p_q^{wB}) - E^{*B}(p_t^{*B}, p_t^{wB})] \\
 & + (p_q - p_q^{wC})[E^{*C}(p_q^{*C}, p_q^{wC}) - E^{*C}(p_t^{*C}, p_t^{wC})] \geq 0
 \end{aligned}$$

(19) clearly holds for a small country: the first line is necessarily 0 because the terms of trade are taken as given ($p_t^{wB} = p_q^{wB} = p^{wB}$); the second line is also 0 as no quota rents arise

³⁹See Bhagwati (1965).

from member countries (i.e. $p_q = p_q^{wB}$). Finally, the third line is positive and it represents exactly the change in tariff revenue from t to the q equilibrium.⁴⁰ Therefore, the following proposition holds true for a small country.

Proposition 2 *For a small country with a competitive market for import quota licenses, $V_q - V_s \geq V_t - V_I$.*

Proof. See the appendix for a direct proof. ■

Note that this proposition holds true also when VER-protection is considered, as it is shown in the appendix. However, this result does not extend to a large country setting due to possible adverse term of trade effects. Recall that the MFN tariff together with the equivalence result which is assumed to hold, imply that $p_s^{wB} = p_I^{wB} = p^w = p_I^{wC} = p_s^{wC}$. Moreover, in equilibrium q , $p_q^{wB} = p_s^{wB}$ and $p_s^{wC} = p_q^{wC}$, while in equilibrium t we have $p_t^{wB}/p_I^{wB} \in [1/\tau_I^{*B}, \tau_I^B]$ and $p_t^{wC} \leq p_I^{wC}$. Like for a small country, the last term in (19) is positive also for a large country because NTBs imply a lower degree of trade diversion resulting in $E^{*C}(p_q^{wC}) \geq E^{*C}(p_t^{wC})$.⁴¹ Supposing $p_t^{wB} \geq p_q^{wB}$, the first term in the first line is positive as imports from B are cheaper in the q equilibrium, but the second term on the same line is negative as trade diversion in equilibrium t is compensated by favorable terms of trade movements. Unless trade of terms gain vis-à-vis non member countries are so strong to offset all other forces, (19) is likely satisfied for a large country who therefore would also gain relative more from a regional integration pursued in presence of NTBs. However, if

⁴⁰Recall that for a small country, after reform q is implemented, imports from C are exactly the quota limit ($E^{*C}(p_q^{*C}, p_q^{wC}) = \overline{M}_s$) and following reform t , $E^{*C}(p_t^{*C}, p_t^{wC}) = 0$ because of complete trade diversion.

⁴¹Recall $p_q^{wC} = p_s^{wC} = p_I^{wC}$ and $p_t^{wC} \leq p_I^{wC}$. Given E^{*C} is an increasing function of world relative price, $E^{*C}(p_q^{wC}) \geq E^{*C}(p_t^{wC})$.

$p_t^{wB} \leq p_q^{wB}$, the first term turns to be negative and the inequality will less likely be satisfied since stronger efficiency gains would be needed. In other words, the stronger the terms of trade gain achieved after the implementation of reform t relative to those attained with reform q , the less likely the welfare outcome of reform q is Pareto superior to the one arising after reform t .

To develop the intuition for this result, it is useful to express the outcome of reform t in its quota-equivalent \bar{m}^e . Indeed, a volume quota of $\bar{m}^e = E^{*C}(p_t^{wC}) \leq \bar{m}_s^C$ set by A on imports from C would deliver the same outcome of reform t induced by the tariff vector τ_I . In equilibrium q , the quota vector results less stringent than in equilibrium t since $\bar{m}_t = (1, \bar{m}^e, 1, 1) \leq \bar{m}_q = (1, \bar{m}_s^C, 1, 1)$. Then, in the thought experiment of a transaction from equilibrium t to equilibrium q , a small country would necessarily gain as this less stringent constraint on imports unambiguously enlarges its consumption possibilities. This result does not extend to large country because of terms of trade effects associated to trade diversion. Recall that p_t^{wB}, p_q^{wB} are the two equilibrium world relative prices in the two reforms. When $p_t^{wB} \leq p_I^{wB} = p_q^{wB}$, A would suffer a terms of trade loss from this hypothetical transaction. Such loss could erode the gains from consumption related to this hypothetical reform since A would end up paying more for all its imports.

Finally, proposition (1) and proposition (2) together suggests that the option for regionalism has turned positive for small and some large countries only in the 90s in presence of quota restricted trade. However, these countries may have correctly anticipated that such valuable option for integration was temporarily restricted by the WTO prescription to convert

all NTBs in tariff-equivalent measures and destined to expire. Therefore, such "tariffication" process - established at the conclusion of the Uruguay round in 1994 and to be completed gradually by 2005 - was easily foreseen by WTO member countries and it may have hurried governments to realize their valuable option. This may be a further contributing factor to the regionalism spurt observed between the 1990 and the first years of 2000.

6 Conclusion

This paper uses a two goods, three countries, general equilibrium model to relate two well known features of trading relations, namely the proliferation and expansion of preferential trade agreements in the 90s to the change in the nature of trade protection occurred over the last thirty years.

It was shown that trade agreements implemented in presence of quota restricted trade naturally implements an explicit-volume preservation rule and, therefore, they result welfare improving. The relevance of this is twofold. First, an explicit volume preservation rule which result cumbersome to implement and very different from the actual WTO prescriptions for PTAs, result naturally implementable and, more importantly, induced by the WTO rules in presence of quota restricted trade. Second, for small and some large countries, preferential trade has become a "positive-dividend" option only in the 90s, when trade was more quota-restricted than in the past, justifying a renewed policy interest for regionalism in the last fifteen years. At the basis of this result is the different mechanism with which a tariff and a quota operate: while a tariff has direct effect on prices, a quota operates on quantities and

affects prices only indirectly. Because a tariff increases the prices of all imports from non-member countries, it may cause union-member to become artificially the lowest cost supplier. Trade is then diverted from non-members toward members of a FTA. Under quota protection this does not occur and a country can still import from the world lowest cost supplier provided that it does not exceed the quota limit. In turn, the pre-agreement volumes of trade are preserved.

Most likely, such benefits associated to regional integration would have expired as soon as countries had complied with the "tariffication" process established by the WTO in 1995 and to be completed in the following ten years. Anticipating such eventuality, governments may have hurried to sign agreements under negotiation.

While this paper contributes to shed some light on the spurt of regionalism in the 90s, it can not provide any answer to the types of agreements signed. Trade policy can not explain whether North-North or South.South or, as more common in the last years, North-South agreements are signed. These themes are certainly worth to be explored in future research. Finally, my model does not allow for strategic interaction of countries and, thus, for plausible retaliation actions. This is because my main interest are the welfare consequences of a trade agreement for a country that trades multilaterally and may employ different instruments of trade policy. In this respect, I follow a long tradition in the literature of piecemeal tariff reform of which regional agreements are just a special kind. This paper shares with this literature also the limit of analyzing only the two polar situations of trade policy: either tariff or non-tariff protection, whereas a framework comprising these instruments together

remains in the research agenda. Clearly, the quota case is useful to understand the effects of RTAs in presence of non-linear tariffs, also defined as a system of tariff-quotas.⁴²

In spite of this limitation, the model correctly predicts that bilateral agreement entailing tariff-reductions can lead to terms of trade effects consistently with empirical findings. More interestingly, given the empirical relevance of VERs, it shows that such effects can also be led by bilateral agreements calling for the elimination of formerly negotiated VERs. Although, the volume preservation outcome is a specificity of country-specific quotas not shared by VERs, the possibility that a welfare improvement will also be the outcome of such agreements abolishing VERs is not compromised.

⁴²Alternatively defined also as tariff-rate quotas. The tariff rate is low under an volume-threshold of imports (on the in-quota quantities), while it is much higher above this threshold (on the out-of-quota quantities). For example, imported car entering under the tariff-quota (up to x cars) are generally charged 10%. Imports entering outside the tariff-quota are charged 80%.

7 Appendix

7.1 Appendix A - Derivation of formula (13)

In general terms, suppose country A undertakes a piecemeal tariff reform such as for example a regional integration agreement. Let me index by 0 the level of a variable in the pre-reform equilibrium and by 1 the level of a variable in the post-reform equilibrium. The budget constraint of country A in the pre-reform and in the post-reform are the following:

$$p(0)C_x(0) + C_y(0) \leq p(0)Q_x(0) + Q_y(0) + R(0) \quad (20)$$

$$p(1)C_x(1) + C_y(1) \leq p(1)Q_x(1) + Q_y(1) + TR(1) \quad (21)$$

where $R(0)$ is the tariff revenue (quota rents) in units of the local export good (y) collected in the pre-reform equilibrium and $TR(1)$ are the total transfers in units of the local export good received in the post-reform equilibrium. The representative consumer of this economy is better off after the implementation of the reform if her utility is at least as high as before the reform. By the weak axiom of revealed preferences (WA), it is enough to show that the old consumption and production bundle are still feasible and affordable in the post-reform equilibrium. Of course, if the transfer compensates the representative consumer in such a way that the old consumption-production bundle is still affordable in the new equilibrium, the WA will hold. Such transfer is the following:

$$TR(1) = [p(1) - p(0)][C_x(0) - Q_x(0)] + R(0) = [p(1) - p(0)]M(0) + R(0) \quad (22)$$

where the consumer receives the old tariff revenue and a subsidy to compensate her for an eventual increase in the relative domestic price of the imported good. It is easy to check that the old consumption-production bundle $(C_x(0), C_y(0), Q_x(0), Q_y(0))$ is still affordable in the new equilibrium by substituting (22) and $((C_x(0), C_y(0), Q_x(0), Q_y(0)))$ in (21):

$$\begin{aligned} p(1)C_x(0) + C_y(0) &\leq p(1)Q_x(0) + Q_y(0) + [p(1) - p(0)][C_x(0) - Q_x(0)] + R(0) \\ &\Leftrightarrow p(0)C_x(0) + C_y(0) \leq p(0)Q_x(0) + Q_y(0) + R(0) \end{aligned}$$

which is always verified as it is (20). Therefore, the consumer is better off. Only it is left to prove is that transfer can be actually financed with the new tariff revenue, so that

$$\begin{aligned} R(1) - TR(1) &= \sum_{i \in \{B, C\}} t^i(1)p^{wi}(1)E^{*i}(1) \\ &\quad - \left\{ [p(1) - p(0)]M(0) + \sum_{i \in \{B, C\}} t^i(0)p^{wi}(0)E^{*i}(0) \right\} \\ &= \sum_{i \in \{B, C\}} t^i(1)p^{wi}(1)E^{*i}(1) \\ &\quad - \sum_{i \in \{B, C\}} [p^{wi}(1)(1 + t^i(1)) - p^{wi}(0)(1 + t^i(0))]E^{*i}(0) + t^i(0)p^{wi}(0)E^{*i}(0) \\ &= \sum_{i \in \{B, C\}} \{ [p^{wi}(0) - p^{wi}(1)]E^{*i}(0) + t^i(1)p^{wi}(1)[E^{*i}(1) - E^{*i}(0)] \} \geq 0 \quad (23) \end{aligned}$$

Note that t^i can also be the implicit tariff associated to an import restriction m^i . The first equality uses (11) and (22), the second equality uses (10) and $p = (1 + t^i)p^{wi}$ and the third equality is just rearranging terms. (23) is exactly (13) where the index 0 is the initial equilibrium s and the index 1 is the post-agreement equilibrium q . Provided (23) holds, the consumer is better off after the reform is implemented and the government budget is balanced.

A similar formula to (23) can also be derived for foreign country i undertaking a similar reform. Using (4), the budget constraint in units of the local export good (x) can be written as:

$$C_x^{*i}(0) + C_y^{*i}(0)/p^{*i}(0) \leq Q_x(0) + Q_y(0)/p^{*i}(0) + R^{*i}(0)$$

$$C_x^{*i}(1) + C_y^{*i}(1)/p^{*i}(1) \leq Q_x(1) + Q_y(1)/p^{*i}(1) + TR^{*i}(1)$$

The following transfer

$$\begin{aligned} TR^{*i}(1) &= \left[\frac{1}{p^{*i}(1)} - \frac{1}{p^{*i}(0)} \right] [C_y^{*i}(0) - Q_y^{*i}(0)] + R^{*i}(0) \\ &= \left[\frac{1}{p^{*i}(1)} - \frac{1}{p^{*i}(0)} \right] M^{*i}(0) + R^{*i}(0) \end{aligned}$$

will make the old consumption-production bundle affordable and therefore the consumer better off. Moreover such transfer is sustainable resulting in a balanced budget by the government provided:

$$R^{*i}(1) - TR^{*i}(1) = \left[\frac{1}{p^{wi}(0)} - \frac{1}{p^{wi}(1)} \right] M^{*i}(0) + \frac{t^{*i}(1)}{p^{wi}(1)} [M^{*i}(1) - M^{*i}(0)] \geq 0 \quad (24)$$

Whenever the last condition holds, foreign country i will be undoubtedly better off following the implementation of the reform.

7.2 Appendix B - Proof of terms of trade preservation under reform q

I shall prove that $p_q^{wB} = p_s^{wB}$.

(8), (10), (9) and $E(p, T) = \sum_{i \in \{B, C\}} M^{*i}(p^{*i}, p^{wi})$ (market y clearing condition) have to

hold in both s and q equilibrium.

Using $E^{*C}(p_q^{wC}) = \bar{m}_s^C = E^{*C}(p_s^{wC})$ and $p_q^{wC} = p_s^{wC}$ and (8),

$$M^{*C}(p_q^{wC}) = M^{*C}(p_s^{wC}) \quad (25)$$

The market x clearing condition (10) implies the following

$$M(p_q, T_q) - E^{*B}(p_q^{*B}, p_q^{wB}) = \bar{m}_s^C = M(p_s, T_s) - \bar{m}_s^B \Leftrightarrow \quad (26)$$

$$M(p_q, T_q) - M(p_s, T_s) = E^{*B}(p_q^{*B}, p_q^{wB}) - \bar{m}_s^B = k$$

The market y clearing condition above together implies that:

$$E(p_s, T_s) - \bar{m}_s^{*B} = M^{*C}(p_s^{wC}) = E(p_q, T_q) - M^{*B}(p_q^{wB}) \Leftrightarrow \quad (27)$$

$$E(p_q, T_q) - E(p_s, T_s) = M^{*B}(p_q^{wB}) - \bar{m}_s^{*B} = H$$

Using (25), (27) in (8), it follows

$$\begin{aligned} \bar{m}_s^{*B} + H &= p_q^{wB} (\bar{m}_s^B + k) \Leftrightarrow \\ p_s^{wB} \bar{m}_s^B + H &= p_q^{wB} (\bar{m}_s^B + k) \\ &\Rightarrow p_s^{wB} = p_q^{wB} \end{aligned}$$

where the second line uses (8) again for equilibrium s . Indeed, when $p_s^{wB} = p_q^{wB}$, then

$H = p_s^{wB} k$ which is reciprocity!

7.3 Appendix C - Reform v reduces \tilde{p}^C and has ambiguous effects on \tilde{p}^B

I shall prove $\tilde{p}_v^C \leq \tilde{p}_o^C$. Recall (8) and (14) to (17) have to hold.

In particular, the assumption of a binding VER (\overline{m}_s^C) even after the reform v is implemented together with (8) imply:

$$\begin{aligned} \frac{M^{*C}(p_o^{*C}, \tilde{p}_o^C)}{\tilde{p}_o^C} &= \overline{m}_s^C = \frac{M^{*C}(p_v^{*C}, \tilde{p}_v^C)}{\tilde{p}_v^C} \\ \Leftrightarrow M^{*C}(p_v^{*C}, \tilde{p}_v^C) &= k M^{*C}(p_o^{*C}, \tilde{p}_o^C) \quad \forall k \geq 0 \\ \tilde{p}_v^C &= k \tilde{p}_o^C \end{aligned}$$

Moreover, reform v changes relative domestic prices lowering relative domestic prices at home because the economic value associated to the artificial scarcity created by a VER is reduced after the reform is implemented. Therefore,

$$\tilde{p}_v^C = p_v \leq p_o = \tilde{p}_o^C \Rightarrow 0 \leq k \leq 1$$

where the equalities between A 's relative domestic price and A 's bilateral terms of trade with C follows from the fact that A is not restraining its exports of good y to C . Therefore, the price an exporter receives at home (p_y) is exactly the price received internationally as paid by the importer (p_y^{*C}).

I shall sketch a proof for $\tilde{p}_v^B \underset{?}{\geq} \tilde{p}_o^B$.

By (8), $\tilde{p}_o^B = \overline{m}_s^{*B} / \overline{m}_s^B$. Since both VERs are assumed to be binding, both M^{*B} and E^{*B} are greater in the new equilibrium. Depending on the relative magnitude of these changes, the bilateral terms of trade with B may result appreciated or depreciated.

7.4 Appendix D - Derivation of (18)

The logic and steps are identical to section 7.1 above. I shall use the same notation as well with the understanding that all variables are now expressed in units of the local import good (good x) at local prices.

$$C_x(0) + C_y(0)/p(0) \leq Q_x(0) + Q_y(0)/p(0) + R(0)$$

$$C_x(1) + C_y(1)/p(1) \leq Q_x(1) + Q_y(1)/p(1) + TR(1)$$

$$TR(1) = [1/p(1) - 1/p(0)][C_y(0) - Q_y(0)] + R(0)$$

$$R(0) = \sum_i [1/\tilde{p}^i(0) - 1/p(0)][C_y^{*i}(0) - Q_y^{*i}(0)]$$

$$R(1) = \sum_i [1/\tilde{p}^i(1) - 1/p(1)][C_y^{*i}(1) - Q_y^{*i}(1)]$$

Adding and subtracting $(\sum_i [1/\tilde{p}^i(1) - 1/p(1)][C_y^{*i}(0) - Q_y^{*i}(0)])$ to $R(1) - TR(1) \geq 0$, gives exactly (18) where $M^{*i}(0) = C_y^{*i}(0) - Q_y^{*i}(0)$.

7.5 Appendix E - Proof of Proposition (2)

Recall A is a small country taking p^{wC}, p^{wB} as given and use the equilibrium result of section (3.1). Under the assumptions of (a) a competitive trade model, (b) of quota license holder operating in a competitive market, (c) of $p^{wC} \leq p^{wB} \leq \tau_I^C p^{wC}$ and d) tariff-equivalent quota, i.e. $\bar{m}_s^C = E^{*C}(p_I^{*C}, p^{wC})$ with an implicit tariff associated to it of τ_I^C , I need to show:

$$V_q - V_s \geq V_t - V_I$$

Proof. Assumptions a), b) and d) imply that $p_I = p_s$. It follows

$$\begin{aligned}
 I_I &= I(p_I, p^{wC}) = p_I Q_x(p_I) + Q_y(p_I) + M(p_I, T_I) t_I^C p^{wC} \\
 &= p_I Q_x(p_I) + Q_y(p_I) + M(p_I, p^{wC})(p_I - p^{wC}) \\
 &= p_s Q_x(p_s) + Q_y(p_s) + \overline{m}_s^C (p_s - p^{wC}) \\
 &= I(p_s, p^{wC}) = I_s
 \end{aligned}$$

where $T_I = p^{wC}$ and $\overline{m}_s^C = E^{*C}(p_I^{*C}, p^{wC}) = M(p_I, p^{wC})$ since A only trades with C in equilibrium I given that the price level and the income in the two trading equilibria are the same, $V_I = V(p_I, I(p_I, p^{wC})) = V(p_s, I(p_s, p^{wC})) = V_s$.

Within the union, producer prices have to be equalized even after the agreement is implemented. Thus,

$$p_q = p_t = p^{wB} \Rightarrow Q_j(p_q) = Q_j(p_t) = Q_j(p^{wB}), \quad j = x, y$$

and

$$\begin{aligned}
 I_q &= I(p_q, T_q) = p^{wB} Q_x(p^{wB}) + Q_y(p^{wB}) + \overline{m}_s^C (p^{wB} - p^{wC}) \\
 &\geq p^{wB} Q_x(p^{wB}) + Q_y(p^{wB}) \\
 &= p_t Q_x(p_t) + Q_y(p_t) \\
 &= I(p_t, p^{wB}) = I_t
 \end{aligned}$$

where the first equality follows from the fact that under quota protection A can still import from the lowest cost supplier C after the regional integration with B up to the

limits imposed by the binding quota. The latter inequality is the direct consequence of trade diversion arising in the tariff case which causes the loss of all tariff revenue after the agreement with B is in place.

By the property of the indirect utility function that is strictly increasing in income, it follows that

$$V_q = V(p_q, I_q) = V(p^{wB}, I(p_q, T_q)) \geq V((p^{wB}, I(p_t, p^{wB}))) = V(p_t, I_t) = V_t$$

■

This result holds true also when VER instead of volume quotas are in place. Terms of trade are unchanged, domestic relative prices are identical to equilibrium q . Thus, $p_o = p_s = p_I$ and $p_v = p_q = p^{wB}$. The only difference is that tariff revenue is accrued by the foreign country.

Proof.

$$\begin{aligned} I_v &= p^{wB}Q_x(p^{wB}) + Q_y(p^{wB}) \\ &= I_t \end{aligned}$$

Then

$$V_q = V(p^{wB}, I_v) = V(p^{wB}, I_t) = V_t$$

Given

$$\begin{aligned}
I_I &= p_I Q_x(p_I) + Q_y(p_I) + M(p_I, T_I) t^C p^{wC} \\
&\geq p_I Q_x(p_I) + Q_y(p_I) \\
&= p_o Q_x(p_o) + Q_y(p_o) \\
&= I_o
\end{aligned}$$

it follows

$$V_o = V(p_o, I_o) \leq V(p_I, I_I) = V_I$$

as $p_o = p_I$. It follows $V_v - V_o \geq V_t - V_I$. ■

References

- [1] Bagwell, Kyle and Staiger, Robert. (1999). An economic theory of GATT. *American Economic Review* 89, 215– 248.
- [2] Bagwell, Kyle and Staiger, Robert. (2002). *The economics of the World Trading System*. Cambridge: MIT Press.
- [3] Bagwell, Kyle and Staiger, Robert. (2004). Multilateral trade negotiations, bilateral opportunism and the rules of GATT/WTO. *Journal of International Economics* 63, 1–29.
- [4] Bhagwati, Jagdish N. (1965). On the Equivalence of Tariffs and Quotas. In: Robert E. Baldwin et al. (eds). *Trade Growth and the Balance of Payments: Essays in Honor of Gottfried Haberler*. Chicago: Rand-McNally, 53-67.
- [5] Bhagwati, Jagdish N. (1993). Regionalism and Multilateralism: An Overview. in De Melo, Jaime and Arvind Panagariya (eds.), *New Dimensions in Regional Integration*, Cambridge: Cambridge University Press.
- [6] Chang, Won and Winters, Alan L. (2002). How Regional Blocs Affect Excluded Countries: The Price Effects of MERCOSUR. *American Economic Review*, forthcoming.
- [7] Chipman, John S. (1987). Compensation Principle, in John Eatwell, Murray Milgate and Peter Newman, eds. *The New Palgrave: A Dictionary of Economics*, vol. 1. London: MacMillan, and New York: Stockton, 524-531.
- [8] De Melo, Jaime, Panagariya, Arvind and Rodrik, Dani. (1993). The New Regionalism: A Country Perspective In, J. de Melo and A. Panagariya (eds.), *New Dimensions in Regional Integration*, London, Cambridge University Press for CEPR, 1993.
- [9] Feenstra, Robert C. (2004). *Advanced International Trade: Theory and Evidence*. Princeton University Press.
- [10] Feenstra, Robert C. and Tracy R. Lewis. (1991b). Negotiated Trade Restrictions with Private Political Pressure. *Quarterly Journal of Economics*, November, 1287-1307.
- [11] Fernandez, Raquel. (1997). Returns to Regionalism: An Evaluation of Non-Traditional Gains from RTAs. *NBER WP* 5970 March.
- [12] Fernandez, Raquel. (1997). Returns to regionalism : an evaluation gains from RTAs. *CEPR Discussion Paper series* 1634.
- [13] Frankel, J.A., Stein, E., and Wei, S.-J. (1996). Regional trading arrangements: natural or supernatural. *American Economic Review* 86(2), 52-56.
- [14] Grinols, Earl L. (1984). The Thorn in the Lion's Paw: Has Britain Paid Too Much for Common Market Membership. *Journal of International Economics*, 16, 271-293.
- [15] Grinols, Earl L. and Kar-yiu Wong. (1991). An Exact Measure of Welfare Change. *Canadian Journal of Economics*, 24(2), 429-449.

- [16] Ju, Jiandong and Kala Krishna. (2000a). Evaluating Trade Reform with Many Consumers. *Canadian Journal of Economics*, 33(3), 787-798.
- [17] Ju, Jiandong and Kala Krishna. (2000b). Welfare and Market Access Effects of Piece-meal Tariff Reform. *Journal of International Economics*, 51, 305-316.
- [18] Kemp, Murray and HenryWan, Jr. (1976). An Elementary Proposition Concerning the Formation of Customs Unions. In: Marry Kemp. *Three Topics in the Theory of International Trade: Distribution, Welfare and Uncertainty*. Amsterdam: North Holland.
- [19] Krishna, Pravin and Panagariya, Arvind. (2002). On Welfare-Enhancing Free Trade Areas. *Journal of International Economics*, 57: 353-367.
- [20] Krueger, Anne O. (1974). The Political Economy of the Rent-Seeking Society. *American Economic Review*, 64(3), June, 291-303.
- [21] Krugman, Paul. (1993). Regionalism versus multilateralism: analytical notes. In: De Melo, J., Panagariya, A. (Eds.). *New Dimensions in Regional Integration*. Cambridge University Press, Cambridge, pp.58-89.
- [22] Laird, S. and Yeat, A. (1990). Trends in Nontariff Barriers of Developed Countries. 1966-1986. *Weltwirtschaftliches Archiv*, 126:299-325.
- [23] Laird, S. and Yeat, A. (1990). *Quantitative Methods for Trade barrier Analysis*. London.
- [24] Markusen, James R., Melvin, James R., Kaempfer, William H., Maskus, Keith E. (1995). *International Trade. Theory and Evidence*. McGraw Hill.
- [25] Summers, Lawrence H. (1991). Regionalism and the World Trading System. In *Policy Implications of Trade and Currency Zones*, Kansas City: Federal Reserve Bank.
- [26] Treffer, Daniel. (2001). The Long and Short of the Canada-U.S. Free Trade Agreement. *NBER Working Paper No. 8293*.
- [27] Viner, J. (1950). *The Customs Union Issue*. New York: Carnegie Endowment for International Peace.
- [28] Vousden, Neil. (1990). *The economics of trade protection*. Cambridge University Press.