Bilateral Trade Liberalization between Asymmetric Countries

Nese Sara*
Department of Economics,
Southern Methodist University,
Dallas, TX 75275-0496.

November 12, 2008

Abstract

The goal of this paper is to identify conditions that give rise to a Free Trade Agreement (FTA) between two countries (called North and South) that are potentially asymmetric with respect to their market sizes and the quality of their products. When market sizes are symmetric, a bilateral FTA does not arise because South (which is assumed to produce the low quality) loses from it and, therefore, rejects it. As the relative market size of North expands, South becomes more willing to sign an FTA, whereas North becomes less so. There exists a range of market sizes, for which an FTA is acceptable to both countries. However, if the asymmetry in market size is too large, an FTA once again fails to arise. As the North-South quality gap increases, the set of market size parameters for which an FTA can be sustained expands. An interesting result is that the globally optimal North-South trade agreement does not call for bilateral free trade. Instead, it calls for free trade in South and a positive tariff in North. Pareto improving unilateral transfers exist, and these expand the market size range supporting an FTA. The direction of the transfer depends on the relative market size. The main results of the paper continue to hold in a repeated game setting when the FTA is self-enforcing, as well as when it is weakly renegotiation-proof.

JEL Classifications: F1, F3, F5.

*E-mail: nsara@smu.edu.

1 Introduction

The failure of the most recent WTO round (i.e., the Doha round) suggests that differences in interests and priorities of diverse economies make it difficult to find common ground while negotiating trade agreements. Yet, large countries, such as the US, have signed several regional and bilateral trade agreements with much smaller countries. There exist several prominent examples of such free trade agreements (FTAs): NAFTA, the US-Chile FTA, the US-Jordan FTA and the EU's free trade agreements with several developing countries. What factors help explain bilateral FTAs between such asymmetric economies? This paper attempts to shed light on this question by focusing on two key dimensions, along which there is substantial variation among the countries locked into such bilateral FTAs: the relative size of their markets and the relative quality of goods produced by them.

A three-stage policy game between North and South is considered, where North is assumed to produce the high quality, and South – the low quality. In the first stage, governments decide whether or not to sign an FTA. In the second stage, countries impose zero tariffs on each other if they commit to an FTA; otherwise, they impose their nationally optimal tariffs. In the final stage, firms compete in prices in each market.[1]

In the model, each government has an incentive to impose an import tariff to protect the domestic firm as well as to extract rents. In general, a country is more willing to sign an agreement when the foreign market is large enough to yield export profits that outweigh the losses that result from eliminating its own tariff. However, this trade-off does not affect the two countries proportionately. More specifically, I show that an FTA arises in equilibrium only if the market size of North is relatively bigger than that of South. An important implication of this is that countries that produce differentiated qualities cannot reach an FTA if their markets are of equal size.

Whether global welfare is greater under an FTA or nationally optimal tariffs depends upon the North-South quality gap. When the quality gap is small and the relative market size of North is large, the global welfare gain from an FTA is negative. More interestingly, when the quality gap is large, the global welfare gain from an FTA is positive regardless of the degree of market size asymmetry. South's welfare gain from an FTA outweighs the loss of North even at substantially asymmetric market sizes.

[1] The model is based on Saggi and Sara (2008) but the key point of departure is that here the focus is on a border measure (i.e. an import tariff) as opposed to a domestic sales tax.

The paper also derives the globally optimal North-South trade agreement; i.e. the pair of tariffs that maximizes their joint welfare. Surprisingly, in this model such an agreement does not call for bilateral free trade. Instead, the globally optimal policy calls for free trade in South, whereas it calls for a positive tariff in North. While North is more willing to enter into the globally optimal trade agreement, rather than an FTA, South is not willing to cooperate unless the export market gains are sufficiently large.

A closely related paper by Park (2000) studies a two-country Ricardian model, where market size is an indicator of a country's ability to manipulate its terms of trade. In his setting, market size asymmetry gives rise to inefficiencies only on the large country's part due to its ability to affect its terms of trade. In my analysis, both countries have an incentive to impose tariffs. As a result, the small country's policy also has efficiency consequences. Indeed, it turns out that the smaller country (i.e. South which is also assumed to produce the low quality) actually creates a relatively larger inefficiency when it pursues its nationally optimal policy than North.

An interesting characteristic of the model is that at least one country always gains from an FTA. This raises the possibility that unilateral transfers can help compensate the losing country. I find that the market size range that supports an FTA expands with transfers. Indeed, the results regarding transfers are also supportive of Lahiri et. al. (2002), who consider a two-stage game in general equilibrium model, where the donor country chooses the income transfer that is tied to tariff reduction of the recipient in the first stage, and tariffs are chosen in the second stage; the authors show that tied aid can be Pareto improving. The way transfers are incorporated into the model is similar to Lahiri et. al.'s when it comes to the timing of events. However, in my setting, transfers do not have strategic effects on the choice of optimal tariffs; they only add to the recipient country's national welfare. Moreover, Lahiri et al. consider unilateral tariff reduction in exchange for transfers; i.e. the donor, in fact, can adjust the level of its trade policy instrument by pulling it above the Nash equilibrium level. In my analysis, a country commits to free trade regardless of whether it is the donor or the recipient. In Park (2000), unless the larger country is compensated by the smaller country with a lump-sum transfer, it does not gain from free trade. In my model, an FTA may also arise with transfer flows going from the larger market (North) to the smaller market (South). For instance, when the market size asymmetry is not significant, it is South that needs compensation for bringing down its tariff barrier since its market access gain is insufficient. In this case, the larger country may need to convince the smaller country to sign the agreement.

In order to capture the enforcement aspect of an FTA, the model is extended to the setting of an infinitely repeated game like Dixit (1987) and Bagwell and Staiger (1990). In this repeated game, a country's deviation from an FTA is punished by the grim trigger strategy of permanent reversion to Nash tariffs. In the game, a self-enforcing FTA arises iff each country's long-term welfare gain from the FTA outweighs its one-shot gain from defecting from the agreement, where the defection results in a permanent breakdown of cooperation. I find that the necessary condition for an FTA to be sustained in equilibrium is that North's market size must lie in the same range as that of the three-stage policy game.[2]

As is well known, if renegotiation is possible, upon a country's defection from the FTA, both countries are strictly better off by forgetting the past and restoring the FTA instead of enforcing the punishment phase. In other words, resuming the FTA equilibrium Pareto dominates permanent reversion to optimal tariffs. In the third part of the paper, I consider a scenario where countries can renegotiate the FTA, and defections are punished by penance strategy instead of permanent reversion to Nash equilibrium. During the punishment phase, the defecting country shows its willingness to restore the FTA by unilaterally cooperating to reduce its tariff to zero, whereas the other country is temporarily allowed to implement its Nash equilibrium tariff as compensation. Using Farrell and Maskin's (1989) concept of weak renegotiation-proofness (WRP), I show that a WRP FTA arises in equilibrium, such that penance with optimal punishment period yields the same incentives for cooperation as permanent reversion to Nash equilibrium. In other words, the results developed for an FTA without renegotiation continue to hold even for an FTA that is WRP.

The next five sections introduce the basic model and present results for the static game. Section 8 solves the model in an infinitely repeated game setting. Unilateral transfers are analyzed in sections 7 and 9, and renegotiation is allowed in section 10. A discussion on autarky punishment strategy is in the appendix.

2 Model

The world is as comprised of two countries (or regions), North and South, where North (N) represents a developed country that is a front-runner in technology, whereas South (S) is a developing country that is lagging behind

[2] In the appendix, it is shown that with a more stringent punishment strategy such as autarky, an FTA can be enforced even for symmetric market sizes.

the technology frontier. Each country has a single domestic firm producing good x. Production of x is vertically differentiated: the domestic firm of North, denoted by h, produces the high quality, while the domestic firm of South produces the low quality (denoted by l). Let s_j be a parameter representing the quality level of x_j, where $s_h \geqslant s_l$. Let r denote the quality ratio where $r \equiv \frac{s_h}{s_l} \geq 1$. Assume that cost of production is symmetric across countries and without loss of generality it is set to zero.

Consumer preferences in each country are represented by the following utility function:

$$u = \theta s_j - p_j \text{ where } j = h, l \tag{1}$$

where, p_j is the price of x_j and θ is a consumer-specific taste parameter. Consumption of quality s_j gives the consumer of valuation θ a gross utility of θs_j. Then, u is the net surplus after spending p_j per unit of consumption. Depending on the value she attaches to consumption of quality s_j, each consumer derives a different utility level. Country i's market size is denoted by μ^i and the taste parameter θ is uniformly distributed over the $\left[0, \mu^i\right]$ interval. Each consumer has to decide whether to purchase the good at all and, if so, what quality to purchase. A consumer does not buy the good and therefore gets zero utility when his taste parameter falls in the $[0, \theta_l^i)$ interval; chooses the low quality if θ falls in the $[\theta_l^i, \theta_h^i)$ interval and chooses the high quality if θ falls in the $[\theta_h^i, \mu^i]$ interval, where

$$\theta_l^i = \frac{p_l^i}{s_l} \text{ and } \theta_h^i = \frac{p_h^i - p_l^i}{s_l(r-1)}. \tag{2}$$

In country i, demand for x_j is as follows:

$$x_j^i(p_l^i, p_h^i) = \begin{cases} \theta_h^i - \theta_l^i = \dfrac{p_h^i - p_l^i}{s_l(r-1)} - \dfrac{p_l^i}{s_l}, \text{ if } j = l. \\ \mu^i - \theta_h^i = \mu^i - \dfrac{p_h^i - p_l^i}{s_l(r-1)}, \text{ if } j = h. \end{cases} \tag{3}$$

A three-stage game is considered in the next section. In the first stage of the game, each government decides whether or not to sign a bilateral Free Trade Agreement (FTA). In the second stage, countries choose their optimal tariffs, given the trade regime. If an FTA has been signed in the first stage, then countries impose zero tariffs on each other's products. If either country is not willing to sign an FTA, both countries discriminate against the foreign product and impose nationally optimal tariffs. In the final stage, firms engage in price competition in each market. I use backwards induction method to solve for the subgame perfect equilibrium of the game.

Throughout the analysis it is assumed that national product markets are segmented.

3 Market Equilibrium under an FTA

Country i imposes a specific tariff on the imported good x_j, denoted by τ_j^i. Given the tariffs, firms maximize their profits in each market. In country i's market, firm j chooses p_j^i in order to maximize $\pi_j^i(p_l^i, p_h^i)$:

$$\underset{p_j^i}{Max}\ \pi_j^i(p_l^i, p_h^i) = (p_j^i - \tau_j^i)x_j^i(p_l^i, p_h^i) = \begin{cases} (p_l^i - \tau_l^i)\left(\dfrac{p_h^i - p_l^i}{s_l(r-1)} - \dfrac{p_l^i}{s_l}\right) \\ (p_h^i - \tau_h^i)\left(\mu^i - \dfrac{p_h^i - p_l^i}{s_l(r-1)}\right) \end{cases} \quad (4)$$

Solving the optimization problem in (4) for domestic and foreign markets gives the following reaction functions at the price competition stage:[3]

$$p_l^i = \frac{p_h^i + \tau_l^i r}{2r} \text{ and } p_h^i = \frac{\mu^i s_l(r-1) + p_l^i + \tau_h^i}{2}. \quad (5)$$

Given country i's tariff τ_j^i, quality parameters and market sizes, the equilibrium prices can be written as:[4]

$$p_l^i = \frac{\mu^i s_l(r-1) + \tau_h^i + 2\tau_l^i r}{4r-1} \text{ and } p_h^i = \frac{r\left(2\mu^i s_l(r-1) + 2\tau_h^i + \tau_l^i\right)}{4r-1}, \quad (6)$$

where $\tau_i^i = 0$ and $\tau_j^i \equiv \tau^i$.

Under an FTA, we have $\tau^i = 0$ and the quality differential is the only cause of price variation. Absent the protective nature of a tariff on the domestic good, price competition between the two qualities becomes more severe.

Lemma 1 summarizes some observations on the prices, output and profits under free trade:

Lemma 1 *Under an FTA, in each market i) the high quality good is more expensive:* $p_h^i(0) > p_l^i(0)$, *ii) the high quality good sells more:* $x_h^i(0) > x_l^i(0)$, *and, as a result, iii) the high quality good yields higher profits:* $\pi_h^i(0) > \pi_l^i(0)$.

[3] Reaction functions slope upward. Firm h is more responsive than firm l to a change in its rival's price: $\frac{\partial p_l^i}{\partial p_h^i} = \frac{1}{2r} < \frac{1}{2} = \frac{\partial p_h^i}{\partial p_l^i}$.

[4] Both prices rise in response to a tariff. Price of the imported good rises more than that of the domestic good: $\frac{\partial p_j^i}{\partial \tau^i} > \frac{\partial p_i^i}{\partial \tau^i} > 0$.

Without the burden of a tariff, the high quality product has a higher profit margin in each market. Under market size symmetry, this affects protectionist incentives of North and South to different degrees. Moreover, an expansion in the market size leads to higher profits for both firms. However, the increase in the profits of the high quality firm is more than that of the low quality firm: protectionist incentives of South grow much stronger when its market size expands, compared to North experiencing such an expansion.[5]

The next section discusses the optimal tariff choice.

4 Nash Equilibrium Tariffs

In the second stage, given the trade policy regime, country i chooses its optimal tariff in order to maximize its national welfare. Country i's national welfare, $w^i(.)$, is the sum of consumer surplus in country i, tariff revenue of the government and profits of firm i in domestic and foreign markets:

$$\underset{\tau^i}{Max}\; w^i(\tau^i, \tau^j) = \sum_j cs^i_j(\tau^i) + \tau^i x^i_j + \pi^i_i(\tau^i) + \pi^j_i(\tau^j). \tag{7}$$

In equation (7) $\tau^i x^i_j$ is the tariff revenue collected by the government of country i, $\pi^i_i(\tau^i)$ and $\pi^j_i(\tau^j)$ are the profits of firm i in the home market and foreign market, respectively. $cs^i_j(\tau^j)$ denotes the consumer surplus. Consumers of the low quality good enjoy a surplus of $cs^i_l(\tau^i)$, whereas the high quality consumers' surplus is $cs^i_h(\tau^i)$:

$$\sum_j cs^i_j(\tau^i) = \int_{\theta^i_l}^{\theta^i_h} (s_l\theta - p^i_l)d\theta + \int_{\theta^i_h}^{\mu^i} (s_h\theta - p^i_h)d\theta, \text{ where} \tag{8}$$

$$cs^i_l(\tau^i) = (\theta^i_h - \theta^i_l)\left[\frac{s_l(\theta^i_h + \theta^i_l)}{2} - p^i_l\right] \text{ and } cs^i_h(\tau^i) = (\mu^i - \theta^i_h)\left[\frac{s_h(\mu^i + \theta^i_h)}{2} - p^i_h\right].$$

As a result of segmented markets, the aggregate welfare of country i can be split into total surplus originating from the domestic market and that from the foreign market. Accordingly, define $\sigma^i(\tau^i)$ as the domestic surplus of country i :

$$\sigma^i(\tau^i) = \sum_j cs^i_j(\tau^i) + \tau^i x^i_j + \pi^i_i(\tau^i). \tag{9}$$

[5] $\frac{\partial \pi^i_l(0)}{\partial \mu^i} > 0$ and $\frac{\partial \pi^i_h(0)}{\partial \mu^i} = 4r\frac{\partial \pi^i_l(0)}{\partial \mu^i}$.

Adding the export profit of country i's domestic firm to equation (9) yields the aggregate welfare of country i :

$$w^i(\tau^i, \tau^j) = \sigma^i(\tau^i) + \pi_i^j(\tau^j). \tag{10}$$

Country i's optimal tariff on firm j, τ^{i*}, under Nash equilibrium is:

$$\tau^{i*} = \frac{\mu^i s_j(r-1)}{3r-2}. \tag{11}$$

Optimal tariffs serve two purposes: to protect the domestic producer against the foreign competitor and to extract rents from the imported good. Without domestic policy instruments, such as domestic taxes or subsidies, imposing a tariff is the only way to achieve both.

Protectionist incentives under market size symmetry: Recall from Lemma 1 that under an FTA the high quality good has a higher profit margin in each market. Under symmetric market sizes, South has a stronger incentive to pursue protectionist policies than North. As a result, South's tariff on the high quality good is greater than North's tariff on the low quality good when market sizes are symmetric:

$$\tau^{S*}\big|_{\frac{\mu^N}{\mu^S}=1} > \tau^{N*}\big|_{\frac{\mu^N}{\mu^S}=1}. \tag{12}$$

Protectionist incentives under market size asymmetry: Inequality (12) does not necessarily hold when market sizes are asymmetric. Under market size asymmetry, the quality gap is not the only determinant of firm profitability. Indeed, under an FTA, a larger domestic market yields higher profits for both firms and makes the foreign firm more attractive for rent extraction. Therefore, each country has a stronger incentive to impose a tariff when its own market expands: $\frac{\partial \tau^{i*}}{\partial \mu^i} > 0$. This also means that stronger protectionist incentive of South under market size symmetry can be balanced by an expansion in North's market. Put differently, if its market size expands, North can be as protectionist as South in that under free trade the low quality firm has higher total profits in the larger North's market. This is due to the fact that low quality exports to a large market can be equally attractive for rent extraction as high quality exports to a small market. In fact, when the market size ratio is equal to the quality ratio, optimal tariffs of the two countries are equal:

$$\tau^{S*}\big|_{\frac{\mu^N}{\mu^S}=r} = \tau^{N*}\big|_{\frac{\mu^N}{\mu^S}=r}. \tag{13}$$

When products become more differentiated, countries impose higher tariffs on the imported goods; *i.e.* $\frac{\partial \tau^{i*}}{\partial r} > 0$. As a result of product differentiation, each firm's share in its export market increases, whereas that in its domestic market decreases. In addition, export profits increase as the quality gap widens and, thus, increases the incentive to extract rents. Therefore, each government imposes a higher tariff on the now more profitable imported product.[6] At the extreme case, when goods are identical, tariffs converge to zero, *i.e.* $\lim\limits_{r\to 1} \tau^{i*} = 0$. This occurs because firm $l's$ quality disadvantage and firm $h's$ price disadvantage fade in their domestic markets. For example, in North's market, even without government intervention domestic high quality good sells for an equal price as the imported good. In other words, firm h deters firm $l's$ entry by charging the same price in North. In this case, there is no need to impose a tariff on the low quality good since it cannot sell in North.

Lemma 2 *Under optimal tariffs, i) the high quality good is more expensive in each market:* $p_h^{i*} \geq p_l^{i*}$, *ii) each good sells more in its domestic market than its export market:* $x_i^{i*} > x_i^{j*}$, *iii) the high quality good always yields higher profits in its domestic market than the low quality good, whereas the low quality good yields higher profits in its domestic market only when the quality differential is small:* $\pi_l^{S*} > \pi_h^{S*}$ *for* $r \in [1, 1.64)$.

As the quality differentiation increases, prices in each market rise: $\frac{\partial p_j^{i*}}{\partial r} > 0, \forall i, j$. Moreover, price of the imported good net of optimal tariff is increasing in the quality gap: $\frac{\partial (p_j^{i*} - \tau^{i*})}{\partial r} > 0,\ i \neq j$. Eliminating the tariff lowers the price of both goods; however, for the imported good, free trade price is greater than the price net of optimal tariff:

$$p_j^i(\tau^{i*}) > p_j^i(0), \forall i, j \text{ and } p_j^i(0) > \left(p_j^i(\tau^{i*}) - \tau^{i*}\right), i \neq j.$$

Sales of a good increase in its export market under an FTA. In contrast, sales of a good in its domestic market are higher under optimal tariffs than under free trade:

$$x_i^i(\tau^{i*}) > x_i^i(0) \text{ and } x_j^i(0) > x_j^i(\tau^{i*}), i \neq j.$$

Higher prices and higher sales under optimal tariffs result in higher profits for a firm in its domestic market, compared to free trade. In contrast, free

[6] The nature of competition in the vertical differentiation model is that an increase in r relaxes price competition. See Shaked and Sutton (1982).

trade price is higher than the price net of tariff, and free trade sales exceed sales under optimal tariffs. Therefore, a firm makes more profits in its export market under free trade:

$$\pi_i^i(\tau^{i*}) > \pi_i^i(0) \text{ and } \pi_j^i(0) > \pi_j^i(\tau^{i*}), i \neq j.$$

Social Efficiency: In order to evaluate the gains and distortions of optimal tariffs and an FTA from a social efficiency standpoint, first we need to consider the way production of goods and their allocation among consumers should be carried out. Once symmetric marginal cost of production is assumed, it is socially efficient to produce and sell only the high quality good. An efficient allocation, therefore, should be such that anyone who can purchase the low quality good should be able to afford the high quality as well: $\theta_l^i = \theta_h^i$. Furthermore, due to zero marginal cost, at socially efficient production level, each market is covered, and consumers of all valuations are served. As discussed later, due to national interests and available policy instruments, countries fail to achieve production and allocative efficiencies.

Lemma 3 *Optimal tariffs cause i) allocative inefficiency: both qualities sell in each market, and ii) underproduction: markets are not covered in any of the countries:*

$$x_j^{i*} > 0 \text{ and } \theta_l^{i*} > 0.$$

Optimal tariffs increase the price differential between the qualities: the high quality sells for more than the low quality. Unlike South, in North's market, despite the tariff imposed on the low quality good, the price of the high quality good is still relatively high.[7] This results in allocative inefficiency: consumers in $\left(\theta_l^{i*}, \theta_h^{i*}\right)$ region buy the low quality good since they cannot afford the high quality. Under the optimal tariffs, prices rise above the marginal cost, and as a result, there are consumers who are not served in each market. In other words, firms underproduce under optimal tariffs.

Lemma 4 *While allocative inefficiency and underproduction still persist under free trade, the underproduction problem in both markets and allocative inefficiency in South are less severe under free trade than under optimal tariffs.*

[7] Absent a subsidy on the high quality good, its price cannot decrease to marginal cost. When marginal costs are equal, with a domestic subsidy, goods h and l would sell for the same price.

More consumers are being served in each market under free trade; *i.e.* $\theta_l^i(\tau^{i*}) > \theta_l^i(0) > 0$, but production is not sufficient to cover the markets. As a result of lower prices, more consumers can afford the goods. Therefore, market coverage increases and underproduction problem is less severe under free trade than under optimal tariffs.

The high quality good is still more expensive than the low quality; therefore, the latter is sold in each market. With free trade, sales of the high quality in South are higher- to some extent, this corrects the inefficiency in allocation in South. Since the opposite holds true in North; i.e. the high quality's sales decrease whereas those of the low quality increase, allocative inefficiency in North deteriorates with free trade.

The next section derives the conditions under which a mutually beneficial FTA arises in equilibrium. For simplicity, let the market sizes of South and North be normalized to 1 and μ respectively; *i.e.* $\mu^S = 1$ and $\mu^N = \mu$.

5 An Endogenous FTA

In the first stage of the game, each country decides whether or not to commit to an FTA, and a binding FTA is signed only if both countries gain from it. In order to see the trade-off each country faces in the first stage of the game, let us look at the components of welfare. Recall from equation (10) that the welfare of country i is the sum of local surplus and the export profit of the domestic firm. An FTA has the opposite effects on the local surplus and the export profits. Consumer surplus is higher under free trade. On the other hand, the government of country i collects tariff revenue, and the profits of domestic firm in the home market are higher with tariff protection. Together these two welfare gains outweigh the loss of consumer surplus; therefore, local surplus is higher under optimal tariffs: *i.e.* $\sigma^i(\tau^{i*}) > \sigma^i(0)$. In contrast, export profits are higher under free trade.

Let Δw^i denote country i's net welfare gain from an FTA:

$$\Delta w^i \equiv w^i(0,0) - w^i(\tau^{i*}, \tau^{j*}). \tag{14}$$

Proposition 1 *A North-South FTA arises in equilibrium iff the relative market size of North lies between two threshold values,* $\overline{\mu}(r)$ *and* $\underline{\mu}(r)$ *where i) North is willing to sign an FTA iff its market size is smaller than* $\overline{\mu}(r)$ *whereas ii) South is willing to sign an FTA iff North's market size is larger than* $\underline{\mu}(r)$*.*

Since the high quality good has a higher profit margin, South has a stronger incentive to impose a tariff on the imported good than North. The

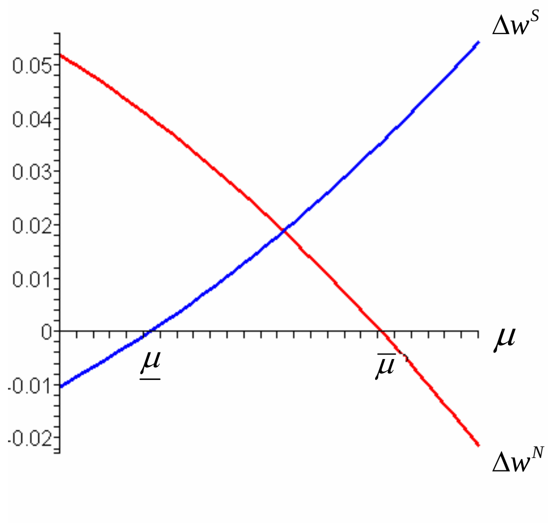

Figure 1: Welfare effects of an FTA: role of market size asymmetry ($r = 1.1$).

quality asymmetry affects protectionist incentives of governments disproportionately. For South, bringing down its optimal tariff is beneficial only when its export market is relatively bigger: only then South can make up for its loss with the increased export revenues. On the other hand, North agrees to free trade even when its export market is not necessarily bigger than its own. However, if its market size becomes relatively too large, North's export gains will not be sufficient to offset the losses resulting from eliminating its optimal tariff. Therefore, the relative market sizes at which both countries commit to an FTA lie between two threshold values, $\underline{\mu}(r)$ and $\overline{\mu}(r)$.

Figure 1 plots net welfare gain functions on the μ axis. Δw^S increases with an increase in μ and attains positive values when μ is above $\underline{\mu}(r)$. In contrast, Δw^N decreases with an increase in μ and North gains from an FTA only at market sizes below $\overline{\mu}(r)$.

Lemma 5 *The difference between the upper and lower critical market sizes is positive at all levels of quality differentiation; hence, the set of market sizes over which an FTA arises in equilibrium is non-empty:*

$$\left(\overline{\mu}(r) - \underline{\mu}(r)\right) > 0, \ \forall r. \tag{15}$$

When the qualities are symmetric, the model turns into a classic Bertrand competition, and countries are indifferent whether to sign an FTA or not:

$\Delta w^i\big|_{r=1} = 0$. At the limit, however, the difference between the critical market sizes is positive, *i.e.* $\lim_{r\to 1}\left(\overline{\mu}(r) - \underline{\mu}(r)\right) = \frac{2-\sqrt{3}}{\sqrt{2}}$.

As the quality gap increases, North is willing to sign an FTA at larger market size values, which were previously beyond the feasible range:

$$\frac{\partial \overline{\mu}(r)}{\partial r} > 0, \ \forall r.$$

The optimal tariff of each country increases with an increase in the quality gap, leading to an increase in export profits and a decrease in domestic surplus. For a given market size, the net welfare gain of North from an FTA also increases, which can be offset if North's market size expands. This is because the expansion in domestic market does not affect export profits, whereas the loss in domestic surplus implied by an FTA increases. On the other hand, an increase in the quality gap can lead to either an increase or a decrease in the net welfare gain of South. When South's net welfare gain falls due to higher quality differentiation, an expansion in μ will offset the loss by increasing only the export profits of South, without altering its domestic surplus (see figure 2). However, when an increase in quality gap leads to a rise in South's net welfare gain, South will be willing to sign an FTA at less asymmetric market sizes than those in the feasible market size range:

$$\frac{\partial \underline{\mu}(r)}{\partial r} > 0 \text{ if } r > 1.14, \ \frac{\partial \underline{\mu}(r)}{\partial r} < 0 \text{ if } r < 1.14. \tag{16}$$

Combining the above findings, we have:

Lemma 6 *As the quality gap widens, the set of market size parameters under which an FTA arises in equilibrium expands:*

$$\frac{\partial\left(\overline{\mu}(r) - \underline{\mu}(r)\right)}{\partial r} > 0. \tag{17}$$

Lemma 7 *An FTA can be sustained in equilibrium iff the market size of North is relatively bigger than the market size of South*; *i.e.* $\mu > 1$.

At symmetric market sizes, North gains and South loses from an FTA: $\Delta w^N\big|_{\mu=1} > 0$ and $\Delta w^S\big|_{\mu=1} < 0$. As the market size of North expands, its gain from an FTA decreases, whereas South's gain increases: $\frac{\partial \Delta w^N}{\partial \mu} < 0$ and $\frac{\partial \Delta w^S}{\partial \mu} > 0$. Furthermore, $\underline{\mu}(r)$ attains its minimum value at $r = 1.14$, where $\underline{\mu}(1.14) = 1.2 > 1$. Therefore, for an FTA to arise in equilibrium, market size of North should be bigger than that of South (see figures 1 and 2).

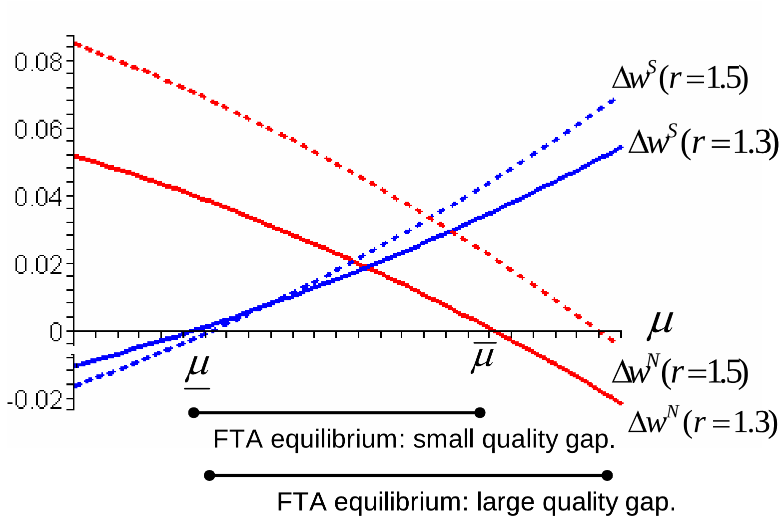

Figure 2: Feasible market size range changes with the quality ratio.

If countries could coordinate their policies to maximize joint welfare, what would be the globally optimal policy? Next section discusses global welfare under nationally and globally optimal tariffs.

6 Globally Optimal Trade Agreement

Global welfare is the sum of national welfare of North and South and is denoted by $ww(.)$:

$$ww(\tau^i, \tau^j) = \sum_i \sigma^i(\tau^i) + \sum_{i \neq j} \pi^i_j(\tau^i). \tag{18}$$

Suppose that countries jointly decide to implement tariffs that maximize global welfare. Is bilateral free trade the globally optimal policy? Surprisingly, it turns out that it is not:

Proposition 2 *Globally optimal policy calls for free trade in South, whereas it calls for a positive tariff in North. Globally optimal tariff vector is denoted by* $\boldsymbol{\tau}^w = (\tau^{Nw}, \tau^{Sw})$:

$$\boldsymbol{\tau}^w = (\tau^{Nw}, \tau^{Sw}) = \left(\frac{\mu(r-1)s_l}{r(4r-3)}, 0 \right) \tag{19}$$

It is globally optimal to limit sales of the low quality good and encourage sales of the high quality good. In South's market, this can be achieved by eliminating the tariff on the high quality good, thus, allowing free trade. This will remedy the distortions to some extent: sales of the high quality good as well as the total sales increase with free trade. In North, imposing the globally optimal tariff on firm l limits its sales and gives an advantage to the high quality good.

Notice that nationally optimal tariff of North is greater than the globally optimal tariff:

$$\tau^{Nw} < \tau^{N*}.$$

The intuition behind the discrepancy is that global optimum takes into account not only the national welfare but profit of the foreign firm as well. Nash equilibrium tariff focuses on increasing the local surplus: foreign firm's profit is important only to the extent that it affects the tariff revenue. As is clear, this leads to a harsher tariff on firm l.

Global optimum calls for a bigger tariff cut in North i) as the relative market size of North expands, and ii) as the North-South quality gap widens:

$$\frac{\partial\left(\tau^{N*}-\tau^{Nw}\right)}{\partial\mu} > 0 \text{ and } \frac{\partial\left(\tau^{N*}-\tau^{Nw}\right)}{\partial r} > 0.$$

First of all, with an expansion in its market size, North becomes more protectionist. Even though the globally optimal tariff of North also increases with an increase in μ, the optimal tariff cut gets bigger with an increase in the relative market size of North. Second, recall that countries impose higher tariffs as the quality gap widens. For a small quality gap, the globally optimal tariff of North increases with an increase in r. However, large quality gap implies higher prices, and it becomes more difficult to provide both qualities. Therefore, the globally optimal tariff of North decreases with an increase in r, at large quality gap.[8] Globally optimal tariff cut increases as the quality gap widens.

Figures 3 and 4 plot the differential between global welfare under globally optimal trade agreement ($ww^w \equiv ww(\tau^{Nw}, \tau^{Sw})$) and nationally optimal tariffs. The welfare differential increases with an increase in μ and r.

A globally optimal trade agreement also has the inefficiencies embodied in an FTA and nationally optimal tariffs. For instance, in North, the globally optimal tariff cannot achieve the market efficiency goal of shutting off firm l from North's market as well as firm h covering the market. First of all, it

[8] $\frac{\partial\tau^{Nw}}{\partial r} < 0$ for $r > 1.5$.

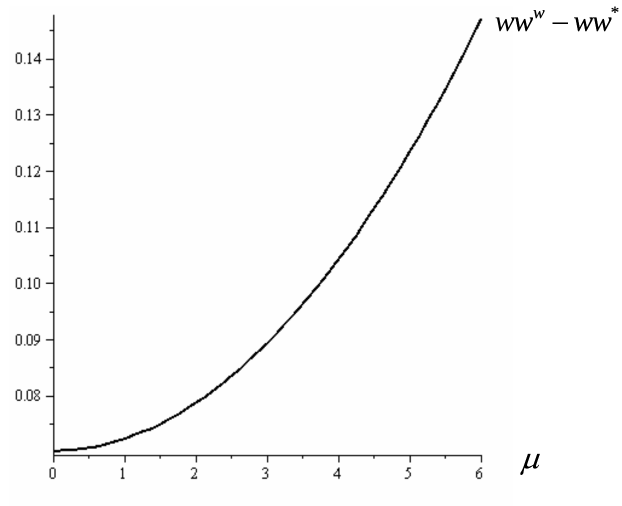

Figure 3: Welfare effect of a globally optimal tariff agreement ($r = 1.5$).

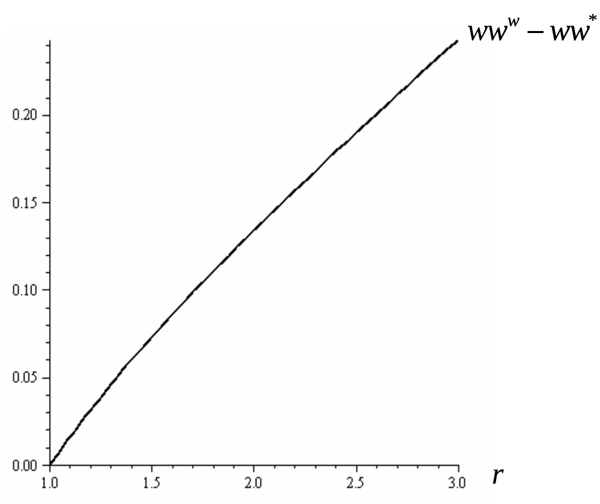

Figure 4: Welfare effect of a globally optimal tariff agreement ($\mu = 1.3$).

is not optimal to impose a tariff which deters firm $l's$ entry to North, since such a tariff would cause market coverage to be even lower and prices to be higher than the global optimum. Second, a tariff that equates the prices of the two qualities is not feasible.[9]

Lemma 8 *South is less willing to cooperate over a globally optimal trade agreement relative to a bilateral FTA, whereas North is more willing to do so:*

$$w^S(0,0) > w^S(\tau^{Sw}, \tau^{Nw}) \text{ and } w^N(0,0) < w^N(\tau^{Nw}, \tau^{Sw}).$$

All else constant, imposing globally optimal tariffs has diverse effects on the incentives of North and South: South is expected to fully liberalize its trade, whereas North is expected only to grant a tariff cut $\left(\tau^{Nw} < \tau^{N*}\right)$. In other words, costs and benefits of global tariff cooperation are distributed disproportionately. The fact that North does not have to fully reciprocate South's trade liberalization, creates an advantageous position for North. Consequently, North gains from cooperating over globally optimal tariffs at even more asymmetric market sizes than those that would result in a gain under an FTA. On the other hand, South is not willing to cooperate unless the export market gains are sufficiently large: for South, only substantially large export markets can justify signing a global optimal trade agreement. As a result, market size asymmetry arises more sharply under global optimum. Cooperation under globally optimal tariffs arises in equilibrium for a more asymmetric market size range, compared to that under a bilateral FTA.[10]

How is the joint welfare affected by the policy regime chosen in the first stage of the game? Let Δww^* denote the net global welfare gain from an FTA:

$$\Delta ww^* \equiv ww(0,0) - ww(\tau^{N*}, \tau^{S*}). \tag{20}$$

Proposition 3 *World welfare improves due to an FTA* ($\Delta ww^* \geq 0$) *if either i)* $r \leq r^*$ *and* $\mu < \mu_g(r)$ *; or ii)* $r > r^*$ *regardless of the difference in*

[9] If both countries would subsidize firm h by $-\mu^i s_l(r-1)$, this subsidy would allow firm h to charge the same price as firm l and fully cover both markets thus achieving social efficiency.

[10] Market sizes that support an FTA and the globally optimal trade agreement do not necessarily overlap. More specifically, for $r < 1.71$ there are no common market sizes for which both an FTA and cooperation under global optimum can arise in equilibrium.

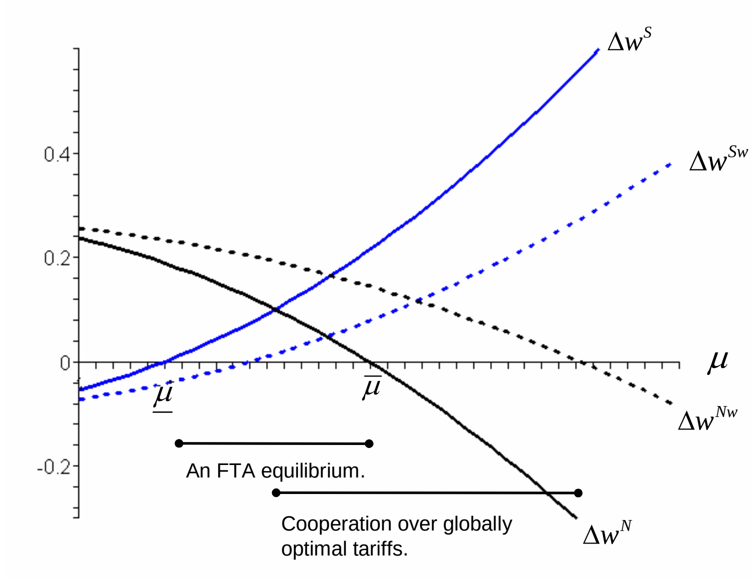

Figure 5: Feasible market size ranges supporting an FTA and the global optimum ($r = 2.5$).

market size:[11]

$$r^* = 1.6404 \text{ and } \mu_g(r) = \sqrt{\frac{r(7r-4)(4r-3)}{9r-4r^2-4}}. \qquad (21)$$

The intuition underlying Proposition 3 is as follows: the only drawback of an FTA is that it increases allocative inefficiency in North. For $r \leq r^*$, if North is a sufficiently large part of the world ($\mu > \mu_g$), its increased allocative inefficiency outweighs the gains from an FTA; thus, global welfare deteriorates under an FTA. In this case, as the market size of North expands, the net global welfare gain decreases; *i.e.* $\frac{\partial(\Delta ww^*)}{\partial \mu} \leq 0$ (see figure 6).

The second part of the proposition highlights the role of quality differentiation. As the quality gap widens, South's distortion of global welfare by its nationally optimal tariff becomes more severe. Therefore, eliminating this distortion by free trade (which is the globally optimal policy for South) contributes much more to global welfare at higher levels of quality gap. Furthermore, above r^* the gap between optimal tariff of North and globally

[11] The critical market size, $\mu_g(r)$, is defined for quality ratios smaller than r^*; *i.e.* $\mu_g^2(r) \geq 0$ iff $r \leq r^*$. And also, $\mu_g(r) > \overline{\mu}(r)$ for $r \leq r^*$.

optimal tariff widens substantially; i.e. $\frac{\partial\left(\tau^{N*}-\tau^{Nw}\right)}{\partial r} > 0$. North's optimal tariff becomes excessively high, intensifies price competition even more and worsens the underproduction problem in North. As a result, an FTA yields higher global welfare than nationally optimal tariffs above r^*. Moreover, an FTA mimics global optimum above the threshold quality gap: above r^*, global welfare under an FTA is positive for all market sizes (see figure 7).

On figure 8 solid lines divide the parameter space into welfare gain and loss areas for North and South ($\Delta w^N = 0$ and $\Delta w^S = 0$ respectively). The area between these two lines shows the parameter values at which both countries are better off under free trade; an FTA arises in equilibrium. The parameter values that lie to the left of the dashed line ($\Delta ww^* = 0$) improve the global welfare under an FTA; however, some of these values cannot support an FTA.[12] Notice that above r^* the net global welfare gain is always positive.

A characteristic of the model is that neither an FTA, nor cooperation over the global optimum can arise in equilibrium at substantially asymmetric market sizes and relatively closer market sizes. This observation raises the following question: since world welfare improves with an FTA for market sizes outside $\left(\underline{\mu}(r), \overline{\mu}(r)\right)$ range, can global welfare gain be redistributed by transfers so that an FTA arises in equilibrium? The role of transfers is discussed next.

7 The Role of Transfers

Suppose that governments can coordinate their policies, so that transfers are available whenever the joint welfare improves with an FTA. As long as one of the countries gains from an FTA, it will be willing to share its gain to induce the other country to sign the agreement. The only motivation for making the transfer is economic: the potential to increase welfare beyond what is possible under optimal tariffs gives one of the countries an incentive to make a transfer and induce the other to sign the FTA. Any strategic effects unilateral transfers may have on the optimal tariffs are ignored; i.e. transfers simply add to the recipient country's welfare.[13]

[12] For instance, under symmetric market sizes joint welfare improves with an FTA $\Delta ww^*(\mu = 1) > 0$. However, countries cannot reach an agreement since South is worse off.

[13] Lahiri et. al (2002) consider a two-stage game where the donor country chooses the income transfer in the first stage, and tariffs are chosen in the second stage; they show that there may be strategic effects on Nash tariffs. More specifically, they show that income transfer may induce the recipient to be more open in its trade policy.

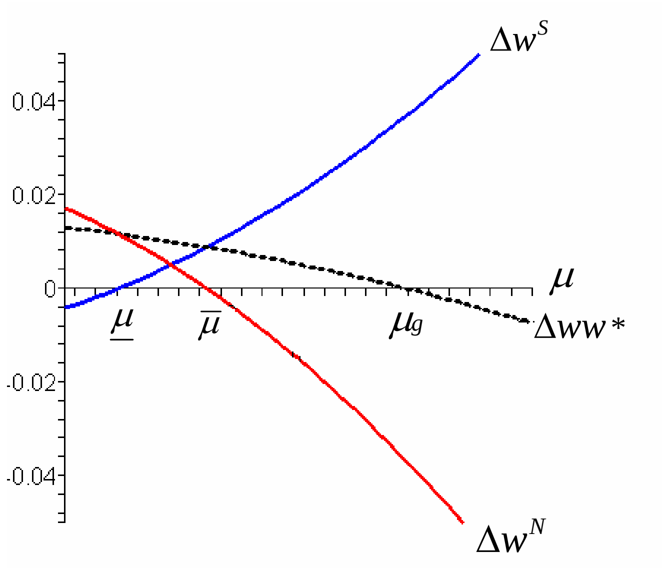

Figure 6: Welfare effects of an FTA - Small quality gap ($r = 1.1$).

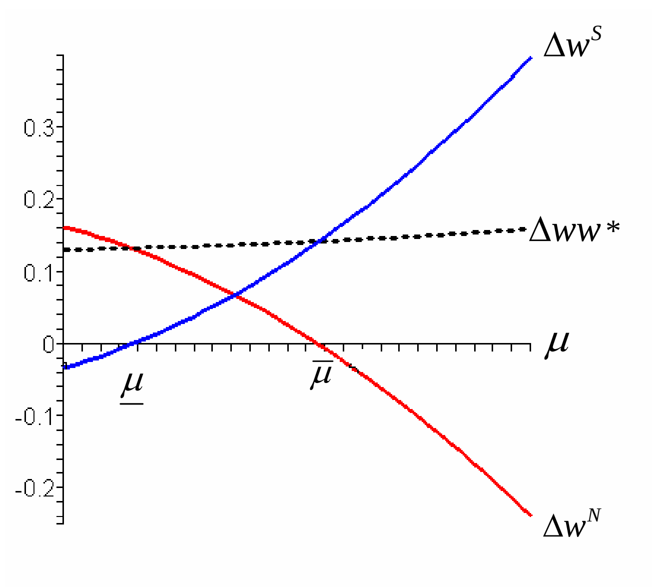

Figure 7: Welfare effects on an FTA - Large quality gap ($r = 2$).

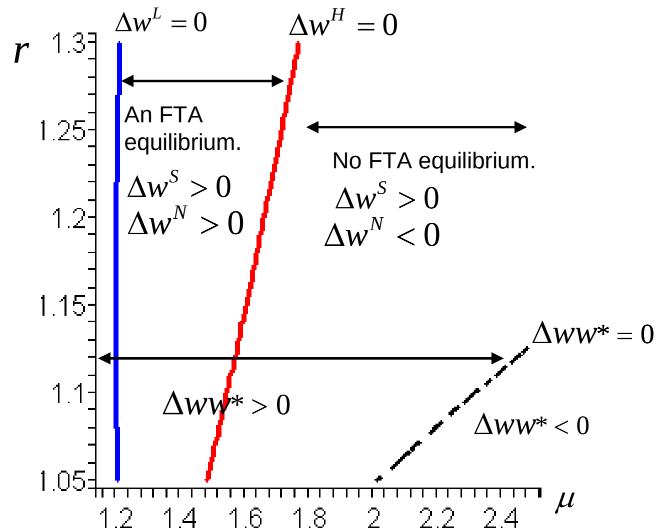

Figure 8: Parameter range for which an FTA arises in equilibrium.

The direction of transfer flows depends on the distribution of net welfare gains from an FTA absent transfers, which in turn depends upon the relative market size. For market sizes below $\underline{\mu}(r)$, North is willing to sign an FTA, whereas South loses from it: when transfer are available, they flow from North to South. Likewise, for market sizes above $\overline{\mu}(r)$ transfers flow from South to North.

Transfers do not always work. In fact, the level of quality differentiation determines the market size range that supports an FTA with transfers. If $r < r^*$, transfers cannot help improving world welfare for market sizes above $\mu_g(r)$: even the maximum transfer from South to North is not sufficient to compensate for the losses of North. However, if $r > r^*$, then world welfare improves with transfers for market sizes above $\mu_g(r)$. Two important observations are as follows:

Lemma 9 *i) Under symmetric market size, an FTA arises in equilibrium with unilateral transfers from North to South, and ii) provided that $r > r^*$, an FTA arises in equilibrium with unilateral transfers from South to North at substantially asymmetric market sizes ($\mu > \mu_g(r)$).*

Note that transfers can be used for cooperation over the globally optimal tariffs. It turns out that the same qualitative nature as under an FTA follows when it comes to the direction of transfer flows. An important exception is that transfers can help cooperate under the global optimum regardless of the degree of r.

Having derived these results, a relevant question to consider is how an FTA is enforced.

8 A Self-enforcing Free Trade Agreement

An FTA is signed only if it is self-enforcing to both countries. For the enforcement of an FTA equilibrium, this section extends the model using infinitely repeated game approach like Dixit (1987) and Bagwell and Staiger (1990). Recalling that an FTA is a long-term commitment, it is reasonable to study it as a repeated interaction between countries. Each country weighs the gains from a one-shot deviation to its Nash equilibrium tariff against the losses from triggering the punishment phase, during which both countries impose optimal tariffs forever.

If country i defects from an FTA, grim trigger punishment strategy of reversion to optimal tariffs (static equilibrium of the previous section) follows forever. When country i defects from an FTA, while country j still commits to a zero tariff, countries' stage game welfare will be:

$$w^i(\tau^{i*}, 0) = \sigma^i(\tau^{i*}) + \pi_i^j(0) \tag{22}$$

$$w^j(0, \tau^{i*}) = \sigma^j(0) + \pi_j^i(\tau^{i*}). \tag{23}$$

Note that for country i, a one-time deviation from an FTA yields greater welfare than both optimal tariffs and free trade. On the other hand, country j gets lower welfare than that under optimal tariffs and free trade. At the end of the first period, punishment phase starts, during which the countries impose their Nash equilibrium tariffs. δ is the common discount factor: $\delta \in (0, 1)$. Country $i's$ and $j's$ discounted sum of welfare is:

$$(1-\delta)w^i(\tau^{i*}, 0) + \delta w^{i*} \tag{24}$$

$$(1-\delta)w^j(0, \tau^{i*}) + \delta w^{j*}. \tag{25}$$

An FTA is incentive compatible for country i, if the discounted sum of FTA welfare, exceeds the discounted sum of welfare from deviation and punishment:

$$w^i(0, 0) \geq (1-\delta)w^i(\tau^{i*}, 0) + \delta w^{i*}. \tag{26}$$

Rewriting the incentive compatibility constraint yields the critical discount factor of country i:

$$\delta \geq \frac{w^i(\tau^{i*},0) - w^i(0,0)}{w^i(\tau^{i*},0) - w^{i*}} \equiv \delta^i(\mu, r). \tag{27}$$

The incentive compatibility constraint for country i holds iff country i values future gains from committing to an FTA sufficiently; i.e. when country i has a sufficiently high discount factor. The critical discount factors for North and South are:

$$\delta^N(\mu, r) = \frac{\mu^2(3r-2)(4r-1)}{2r(2r-1)(10r-7)}, \tag{28}$$

$$\delta^S(\mu, r) = \frac{r(3r-2)(4r-1)}{2\mu^2(2r-1)(4r-3)}. \tag{29}$$

Lemma 10 *A self-enforcing FTA arises in equilibrium iff* $\delta \geq \delta^i(\mu, r)$, $\forall i$.

An FTA is subgame perfect equilibrium, if both countries' incentive compatibility constraints are satisfied. Analyzing (28) and (29), we see that for the feasible range of δ values, the critical discount factors never attain a value of 0. However, there exist δ values for which an FTA fails to be sustained in equilibrium:

$$\delta^N = 1 \text{ if } \mu \geq \overline{\mu}(r), \text{ and } \delta^S = 1 \text{ if } \underline{\mu}(r) \geq \mu. \tag{30}$$

Proposition 4 *Under punishment of permanent reversion to optimal tariffs, an FTA may be sustained in equilibrium if North's market size lies between* $\overline{\mu}(r)$ *and* $\underline{\mu}(r)$, *since i) North defects from an FTA if its market size is at least* $\overline{\mu}(r)$, *and ii) South defects from an FTA if North's market size is less than or equal to* $\underline{\mu}(r)$:

$$\delta^i < 1, \ \forall i \ \textit{iff} \ \overline{\mu}(r) > \mu > \underline{\mu}(r).$$

Proposition 4 holds true because the threat of reversion to Nash equilibrium forever constrains the FTA by the same market size range as that of the static game. If North's market size lies in the $\left(\underline{\mu}(r), \overline{\mu}(r)\right)$ interval, the critical discount factors cannot attain a value of 1. Lemma 1 continues to hold; *i.e.* the set of market sizes that may support an FTA is nonempty: $\left(\overline{\mu}(r) - \underline{\mu}(r)\right) > 0$, $\forall r$.[14]

[14] Under a more severe punishment strategy, an FTA may be self-enforcing. In the appendix, it is shown that with autarky punishment an FTA arises in equilibrium even when the countries have symmetric market sizes.

Let $\mu^*(r)$ be the market size that equates the critical discount factors: $\delta^N(\mu^*) = \delta^S(\mu^*)$. If North's market size is between $\mu^*(r)$ and $\overline{\mu}(r)$, then (26) binds for North. If it is between $\underline{\mu}(r)$ and $\mu^*(r)$, (26) binds for South.

$$\begin{aligned} \delta^N(\mu,r) &> \delta^S(\mu,r), \text{ if } \mu > \mu^*(r), \\ \delta^N(\mu,r) &< \delta^S(\mu,r), \text{ if } \mu < \mu^*(r), \\ \delta^N(\mu,r) &= \delta^S(\mu,r), \text{ if } \mu = \mu^*(r). \end{aligned} \tag{31}$$

If the North's market size expands, the $(0,\delta^N)$ interval widens. The potential gains of North from committing to an FTA get smaller when μ increases: this leads to a greater decrease in domestic surplus, whereas export profit remains unchanged. Therefore, North becomes indifferent whether to defect from or commit to an FTA at a higher patience level: $\frac{\partial\delta^N}{\partial\mu} > 0$. Unlike North, South's critical discount factor decreases if μ expands: $\frac{\partial\delta^S}{\partial\mu} < 0$. As its export market gets bigger, South's potential gain from an FTA increases, making it indifferent whether to defect from or commit to an FTA at a lower patience level.

In the infinitely repeated game, transfers come into the picture whenever an FTA is self-enforcing to one country but not to the other. The role of transfers in enforcing an FTA is discussed next.

9 Is Free Trade Sustainable with Transfers

Once unilateral transfers are available, a country's gain from a one-time deviation and the associated loss from breaking the FTA cooperation forever will not balance anymore. Thus, the incentive compatibility constraint changes. This section analyzes how transfers affect the sustainability of an FTA.

Let T^i denote the transfer from country j to country i. As before, incentive compatibility constraints yield the critical discount factors. $\delta^i(\mu,r,T^i)$ is the discount factor at which country i is indifferent whether to defect after receiving the transfer T^i or to commit to an FTA. At $\delta^j(\mu,r,T^i)$ country j is indifferent whether to defect without making any transfers or to commit to an FTA under which country j transfers to country i:

$$\delta^i(\mu,r,T^i) = \frac{w^i(\tau^{i*},0) - w^i(0,0)}{w^i(\tau^{i*},0) - w^{i*} + T^i} \text{ and } \delta^j(\mu,r,T^i) = \frac{w^j(\tau^{j*},0) - w^j(0,0) + T^i}{w^j(\tau^{j*},0) - w^{j*}}. \tag{32}$$

Lemma 11 *An FTA with Pareto improving transfers arises in equilibrium iff* $\delta \geq \delta^i(\mu,r,T^j)$, $\forall i,j$.

Consider the case that $\delta \in (\delta^j, \delta^i)$: an FTA is incentive compatible to country j but not to country i. A transfer from country j to country i, decreases country i's incentive to defect from an FTA, whereas it leads to an increase in country j's incentive to defect:

$$\delta^i(\mu, r, T^i)\big|_{T^i>0} \leq \delta^i(\mu, r, T^i)\big|_{T^i=0} \text{ and } \delta^j(\mu, r, T^i)\big|_{T^i>0} \geq \delta^j(\mu, r, T^i)\big|_{T^i=0}. \tag{33}$$

The critical discount factors of North and South attain a value of 1, at $\overline{\mu}(r, T^i)$ and $\underline{\mu}(r, T^i)$, respectively. For market sizes outside the $(\underline{\mu}(r, T^i), \overline{\mu}(r, T^i))$ range, an FTA fails to arise in equilibrium:

$$\delta^N(\mu, r, T^i) = 1 \text{ if } \mu \geq \overline{\mu}(r, T^i) \text{ and } \delta^S(\mu, r, T^i) = 1 \text{ if } \mu \leq \underline{\mu}(r, T^i). \tag{34}$$

Since transfers alter the incentives to defect from an FTA, critical market sizes also change. When North is the recipient, it is more willing to sign a bilateral FTA; therefore, its critical market size shifts beyond $\overline{\mu}(r)$ to $\overline{\mu}(r, T^i)$. Since making a transfer affects South's incentive in the opposite direction, its critical market size is larger than that without a transfer. When South is the recipient, it is willing to sign a bilateral FTA at market sizes even smaller than $\underline{\mu}(r)$. North is less willing to sign an FTA when it makes a transfer: its critical market size is smaller than that without a transfer.

Lemma 12 *The market size range supporting an FTA expands with unilateral transfers: i) when transfers flow from South to North, an FTA can arise in equilibrium for more asymmetric market sizes, ii) when transfers flow from North to South, an FTA can arise in equilibrium for relatively closer market sizes:*[15]

$$\overline{\mu}(r, T^N) \geq \overline{\mu}(r) \geq \underline{\mu}(r) \geq \underline{\mu}(r, T^S). \tag{35}$$

What is the direction of transfer flows? Recall that critical discount factors are equal at $\mu^*(r)$. If an FTA is incentive compatible to South but not to North (above $\mu^*(r)$), South may find it profitable to make a transfer to North. Below $\mu^*(r)$, we have the opposite case.

Proposition 5 *The relative market size of North determines the direction of transfers. Transfers flow i) from South to North at market sizes above $\mu^*(r)$, and ii) from North to South at market sizes below $\mu^*(r)$.*

[15] $\overline{\mu}(r, T^N) \geq \overline{\mu}(r) \geq \overline{\mu}(r, T^S)$ and $\underline{\mu}(r, T^N) \geq \underline{\mu}(r) \geq \underline{\mu}(r, T^S)$.

Let $\underline{T}^i$ denote the minimum transfer from country j acceptable to country i. At $\underline{T}^i$ country i is indifferent whether to defect upon receiving the transfer or to commit to an FTA. $\overline{T^i}$ denotes the maximum amount country j is willing to transfer to country i. At $\overline{T^i}$ country j is indifferent whether to defect without making any transfers or to commit to an FTA with transfers. By definition, at $\underline{T}^i$ and $\overline{T^i}$, the incentive compatibility constraints hold with equality:

$$\underline{T}^i(\mu, r, s_l, \delta) = \frac{1}{\delta}(w^{iD} - w^i(0,0)). \tag{36}$$

$$\overline{T^i}(\mu, r, s_l, \delta) = w^j(0,0) - w^{jD}. \tag{37}$$

An FTA equilibrium with transfers satisfies the following conditions: *i)* country i must be willing to accept the transfer: $T^i \geq \underline{T}^i \geq 0$; *ii)* country j must be willing to make a transfer: $\overline{T^i} \geq T^i \geq 0$, and *iii)* the maximum transfer offered by country j should be at least equal to the minimum amount acceptable to country $i : \overline{T^i} \geq \underline{T}^i$. Let $\Delta T^i(\mu, r, s_l, \delta)$ be the difference between minimum transfer acceptable to country i and maximum transfer offered by country j:

$$\Delta T^i \equiv \overline{T^i}(\mu, r, s_l, \delta) - \underline{T}^i(\mu, r, s_l, \delta). \tag{38}$$

Lemma 13 *The set of parameters for which an FTA with Pareto improving transfers arises in equilibrium is non-empty.*

Figure 9 plots $\Delta T^N = 0$ and $\Delta T^S = 0$ on (δ, μ) space and superimposes the critical discount factors. An FTA naturally arises in equilibrium for parameter values in the $(\underline{\mu}(r), \overline{\mu}(r))$ interval and above the critical discount factors. To the right of $\mu^*(r)$, an FTA is incentive compatible to South but not to North whenever δ lies in the $\left(\delta^S, \delta^N\right)$ interval. The shaded area to the right of $\mu^*(r)$ shows parameter values for which an FTA with transfers from South to North arises in equilibrium; i.e. $\Delta T^N > 0$ and $\delta \in \left(\delta^S, \delta^N\right)$. To the left of $\mu^*(r)$, an FTA is incentive compatible to North but not to South whenever δ lies in the $\left(\delta^N, \delta^S\right)$ interval. The shaded area to the left of $\mu^*(r)$ shows parameter values for which an FTA with transfers from North to South arises in equilibrium; i.e. $\Delta T^S > 0$ and $\delta \in \left(\delta^N, \delta^S\right)$.[16]

[16] If permanent reversion to autarky was chosen as punishment strategy instead, $\underline{T}^i$ would be lower and $\overline{T}^i$ would be higher; therefore, ΔT^i would increase. In other words, a harsher punishment would increase both willingness to accept and willingness to make a transfer, hence, making it easier to sustain an FTA equilibrium. Park (2000) argues that autarky punishment affects countries of asymmetric size in a direction favorable to only the small country.

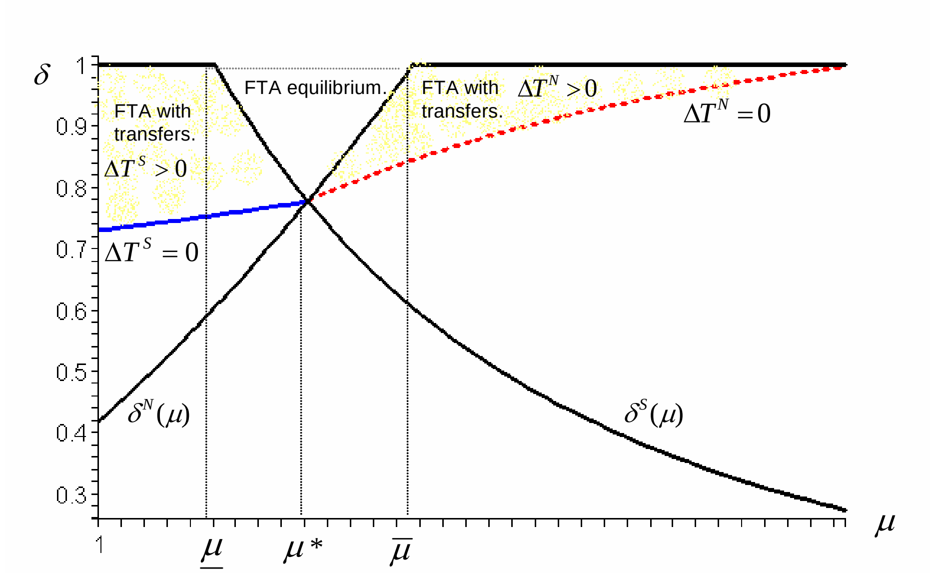

Figure 9: The set of parameters for which an FTA with Pareto improving transfers arises.

10 Weakly renegotiation-proof FTA

If renegotiation is allowed upon a country's defection, both countries will be strictly better off by forgetting the past and restoring the FTA equilibrium instead of enforcing the punishment phase: $w^i(0,0) \geq w^{i*}$. Resuming the FTA equilibrium Pareto dominates permanent reversion to the optimal tariffs. As is clear, the possibility of renegotiation eliminates the effectiveness of the threat of punishment. I use Farrell and Maskin's (1989) concept of weak renegotiation-proofness (WRP) with the following asymmetric punishment strategy: during penance punishment phase, the defecting country shows its willingness to restore the FTA by unilaterally setting zero tariff, whereas the other country is compensated by temporarily implementing its Nash equilibrium tariff. Indeed, during the punishment phase, the punishing country acts like the deviating country and the deviator willingly does not best respond; thus, payoffs are exchanged.[17]

The incentive compatibility constraint for country i can be written as

[17]For different definitions of renegotiation-proofness, see Pearce(1987) and Abreu, Pearce, Stachetti (1993). Also see Fudenberg and Tirole (1991).

follows:

$$w^i(0,0) \geq (1-\delta)\left(w^i(0,\tau^{i*}) + \sum_{t=1}^{n} \delta^t w^i(\tau^{j*},0) + \sum_{t=n+1}^{\infty} \delta^t w^i(0,0)\right). \quad (39)$$

For a punishment strategy to be WRP, the following should be satisfied: *i)* the deviating country should agree to the punishment, i.e. should not find it profitable to walk away from the FTA by imposing its optimal tariff forever:

$$(1-\delta)\left(\sum_{t=1}^{n} \delta^t w^i(\tau^{j*},0) + \sum_{t=n+1}^{\infty} \delta^t w^i(0,0)\right) \geq \delta w^{i*}, \quad (40)$$

ii) the punishing country should be willing to carry out the punishment instead of leaving the FTA by imposing its optimal tariff forever:

$$(1-\delta)\left(\sum_{t=1}^{n} \delta^t w^j(0,\tau^{j*}) + \sum_{t=n+1}^{\infty} \delta^t w^j(0,0)\right) \geq \delta w^{j*}. \quad (41)$$

Combining (40) and (41) gives:

$$\frac{w^j(0,\tau^{j*}) - w^{j*}}{w^j(0,\tau^{j*}) - w^j(0,0)} \geq \delta^n \geq \frac{w^{i*} - w^i(\tau^{j*},0)}{w^i(0,0) - w^i(\tau^{j*},0)}. \quad (42)$$

And finally, *iii)* there should be no alternative punishment strategy that Pareto dominates penance punishment. Obviously, the defecting country would be better off by resuming the FTA without going through the punishment. However, given a chance to unilaterally implement its Nash equilibrium tariff, the punishing country finds it profitable to carry out the punishment instead of resuming the FTA:[18]

$$(1-\delta)\left(\sum_{t=1}^{n} \delta^t w^j(0,\tau^{j*}) + \sum_{t=n+1}^{\infty} \delta^t w^j(0,0)\right) \geq \delta w^j(0,0). \quad (43)$$

Let n^* be the punishment period for which (40) holds with equality: $\delta^{n^*} = \left(\frac{w^{i*}-w^i(\tau^{j*},0)}{w^i(0,0)-w^i(\tau^{j*},0)}\right)$. For n^*, the deviating country's welfare during the punishment phase is what it could receive with permanent reversion to

[18] The FTA equilibrium brings welfare that is higher than any punishment; therefore, other punishment strategies (such as permanent reversion to autarky) cannot be weakly renegotiation-proof. This also means that permanent reversion to optimal tariffs is subgame perfect but not WRP.

its optimal tariff. This yields the minimum welfare deviating country can achieve even if it steps out of the FTA forever:[19]

Country i's loss from cooperation when country j defects from the FTA, is equal to its loss from the punishment when country i itself becomes the deviator: $w^i(0,0) - w^i(\tau^{j*},0) = w^i(0,\tau^{i*}) - w^{i*}$. Similarly, country i's gain from best responding to j's defection is equal to its gain from defection when roles are reversed: $w^{i*} - w^i(\tau^{j*},0) = w^i(0,\tau^{i*}) - w^i(0,0)$. Using these equalities, critical discount factor of the deviating country can be written as a function of the actual discount factor and optimal punishment period:

$$\delta^i(\mu, r) = \delta^{n^*}(\mu, r). \tag{44}$$

Proposition 6 *A WRP and self-enforcing FTA arises in equilibrium with n^* periods of penance punishment:*

$$n^*(\mu, r, \delta) = \frac{\ln\left(\delta^i\right)}{\ln\left(\delta\right)}. \tag{45}$$

Penance punishment that lasts for n^* periods, yields the same critical discount factors as permanent reversion to optimal tariffs. An FTA with such punishment is WRP. If countries can renegotiate the agreement, the same set of common discount factors still satisfies the incentive compatibility constraints. This implies that the results developed for an FTA without renegotiation continue to hold for a WRP FTA.[20]

Note that at the limit when country i is indifferent whether to defect from or to commit to an FTA, reversing the roles between the countries by only one period is enough to create the same incentives as a permanent reversion to optimal tariffs: $n^*(\mu, r, \delta^{i*}) = 1$. Furthermore, as the incentive to deviate increases, optimal punishment period extends: $\frac{\partial n^*(\mu,r,\delta)}{\partial \delta} < 0$.

11 Conclusion

The paper asks when and why an FTA is incentive compatible for countries of potentially asymmetric market sizes that produce vertically differentiated

[19] In deriving the endogenous punishment period, I followed Limao and Saggi (2007).

[20] A WRP FTA equilibrium arises due to the assumption of perfect foresight. If we consider a trade agreement as an incomplete contract, then countries may need to renegotiate after experiencing a shock. Moreover, this finding is not robust to different definitions of renegotiation-proofness. For example, Ludema (2002), using renegotiation-proofness in the sense of Pearce (1987), analyzes the role of DSU as a renegotiation platform. He shows that free trade would be achieved when countries cannot renegotiate, but it does not arise in equilibrium in the presence of DSU.

goods. In order to address this question, a North-South bilateral trade policy game is considered with the following timing of events: in the first stage, each government decides whether or not to sign an FTA. In the second stage, countries choose their optimal tariffs, given the trade regime. In the final stage, given tariffs, firms engage in price competition in each market.

A bilateral FTA benefits a country as long as its export market gain is large enough to exceed the loss resulting from eliminating its tariff. When countries are of equal size South is more protectionist than North, and is not willing to eliminate its tariff. Therefore, an FTA does not arise in equilibrium. An expansion in North's market size makes South more willing to sign an FTA, whereas it decreases North's willingness to do so. An FTA arises in equilibrium only if the market size of North is relatively bigger than that of South.

The paper compares global welfare under an FTA and nationally optimal tariffs. If North-South quality gap is large, global welfare is always greater under an FTA. However, if the quality gap is small, global welfare is greater under an FTA only if the relative market sizes of North and South are not substantially asymmetric. It is also shown that the globally optimal North-South trade agreement is not a bilateral FTA; rather, it is an agreement that calls for free trade in South and a positive tariff in North.

As long as global welfare improves under an FTA, unilateral transfers may help compensate the country that loses from it. It is shown that Pareto improving transfers exist, and these expand the market size range supporting an FTA. In the model, direction of transfers depends on the relative market sizes of the countries.

Solving the model as an infinitely repeated game with grim trigger punishment strategy shows that a self-enforcing FTA may arise in equilibrium only if North's market size lies in the same range as that of the three-stage policy game. Finally, the paper asks whether an FTA arises in equilibrium when countries can renegotiate the agreement and the model generates a weakly renegotiation-proof FTA, for which the main findings continue to hold.

Using domestic production subsidies would alter some of the results and such a scenario would reflect the linkage between a country's domestic and trade policies. It is also worthwhile to add a quality investment stage to check for the robustness of the results against leapfrogging. Also, the model does not capture the effects of an FTA in a multilateral setting, such as trade creation, trade diversion effects and coalitional dynamics. A richer setting including a third country may address these issues.

12 Appendix

12.1 Autarky punishment

Let $\delta^{ia}(\mu, r)$ denote the critical discount factor for country i under autarky punishment. The incentive compatibility constraint for country i holds if the common discount factor is at least equal to the critical discount factors.

With autarky punishment, $\delta^{ia}(\mu, r)$ is strictly less than 1 for all market sizes. For a sufficiently high δ, an FTA can be self-enforcing for market sizes that were not in the feasible range under the less strict punishment strategies. In fact, we have the following ranking of the critical discount factors:

$$\delta^i > \delta^{ia} \text{ and } \delta^{ia} < 1, \ \forall i. \tag{46}$$

Lemma 14 *An FTA arises in equilibrium under autarky punishment when market size is symmetric across countries.*

The more strict a punishment strategy is, the easier it is to sustain FTA equilibrium. This results from the impact of punishment on the cost of defection: among different punishment strategies the initial gain from defection is the same; however, the cost of punishment increases with the severity of punishment. It is easier to enforce an FTA when defections are punished by a reversion to autarky rather than a reversion to optimal tariffs.

Park (2000) argues that severity of punishment affects countries with asymmetric size differently. Unlike Park (2000), here the severity of punishment affects both countries in the same direction: North is willing to sign an FTA for smaller market sizes, and South for more asymmetric market sizes with a more stringent punishment.

12.2 Equations

Nash equilibrium tariffs

$$p_l^{S*} = \frac{2\mu^S s_l(r-1)(2r-1)}{(3r-2)(4r-1)} \text{ and } p_h^{S*} = 2rp_l^{S*} \tag{47}$$

$$p_l^{N*} = \frac{\mu^N s_l(r-1)(5r-2)}{(3r-2)(4r-1)} \text{ and } p_h^{N*} = \frac{3r(2r-1)}{(5r-2)} p_l^{N*} \tag{48}$$

$$x_l^{S*} = \frac{2\mu^S r(2r-1)}{(3r-2)(4r-1)} \text{ and } x_h^{S*} = \frac{(4r-3)}{2(2r-1)} x_l^{S*} \tag{49}$$

$$x_l^{N*} = \frac{\mu^N r(r-1)}{(3r-2)(4r-1)} \text{ and } x_h^{N*} = \frac{3(2r-1)}{(r-1)} x_l^{N*} \tag{50}$$

$$\pi_l^{N*} = \frac{\left(\mu^N\right)^2 s_l r(r-1)^3}{(3r-2)^2\,(4r-1)^2} \text{ and } \pi_h^{N*} = \frac{9r\,(2r-1)^2}{(r-1)^2} \pi_l^{N*} \tag{51}$$

$$\pi_l^{S*} = \frac{4\left(\mu^S\right)^2 s_l r(r-1)\,(2r-1)^2}{(3r-2)^2(4r-1)^2} \text{ and } \pi_h^{S*} = \frac{r(4r-3)^2}{4(2r-1)^2} \pi_l^{*S} \tag{52}$$

$$\theta_l^{S*} = \frac{2\mu^S (r-1)(2r-1)}{(3r-2)\,(4r-1)} \text{ and } \theta_l^{N*} = \frac{\mu^N (r-1)(5r-2)}{(3r-2)\,(4r-1)}. \tag{53}$$

Market equilibrium under an FTA

$$p_l^i\,(0) = \frac{\mu^i s_l (r-1)}{(4r-1)} \text{ and } p_h^i\,(0) = 2r p_l^i\,(0)\,. \tag{54}$$

$$x_l^i\,(0) = \frac{\mu^i r}{(4r-1)} \text{ and } x_h^i\,(0) = 2 x_l^i\,(0)\,. \tag{55}$$

$$\pi_l^i(0) = \frac{\left(\mu^i\right)^2 s_l r(r-1)}{(4r-1)^2} \text{ and } \pi_h^i(0) = 4r\pi_l^i(0). \tag{56}$$

$$\theta_l^i(0) = \frac{\mu^i (r-1)}{4r-1} > 0. \tag{57}$$

$$\frac{\partial p_j^i(0)}{\partial r} > 0 \text{ and } \frac{\partial x_j^i(0)}{\partial r} < 0. \tag{58}$$

An endogenous FTA

$$\overline{\mu}(r, s_l) = \sqrt{\frac{2r(2r-1)(10r-7)}{(3r-2)\,(4r-1)}} > \underline{\mu}(r) = \sqrt{\frac{r\,(3r-2)\,(4r-1)}{2\,(2r-1)\,(4r-3)}}. \tag{59}$$

Globally optimal tariffs

$$x_l^{Nw} = \frac{\mu(r-1)}{(4r-3)} \text{ and } x_h^{Nw} = \frac{(2r-1)}{(r-1)} x_l^{Nw}. \tag{60}$$

$$p_l^{Nw} = \frac{\mu(r-1)s_l}{(4r-3)} \text{ and } p_h^{Nw} = (2r-1)p_l^{Nw}. \tag{61}$$

A Self-enforcing FTA

$$\mu^* = \left(\frac{r^2(10r-7)}{4r-3}\right)^{\frac{1}{4}}. \tag{62}$$

A Self-enforcing FTA with Transfers

$$\delta^i(\mu, r, T^i) = \frac{w^{id} - w^i(0,0)}{w^{id} - w^{i*} + T^i} \text{ and } \delta^j(\mu, r, T^i) = \frac{w^{jd} - w^j(0,0) + T^i}{w^{id} - w^{i*}}. \tag{63}$$

$$A = \frac{2(3r-2)(4r-1)}{s_l r(r-1)} \tag{64}$$

$$\delta^N(\mu, r, T^N) = \left(\frac{1}{\delta^N(\mu, r)} + T^N \frac{A}{\mu^2}\right)^{-1} \text{ and } \delta^S(\mu, r, T^N) = \delta^S(\mu, r)(1 + T^N \frac{A}{r}). \tag{65}$$

$$\delta^N(\mu, r, T^N) = 1 \text{ if } \mu \geq \overline{\mu}(r, T^N) = \sqrt{\overline{\mu}^2 \left(1 + T^N \frac{A}{\overline{\mu}^2}\right)} \tag{66}$$

$$\delta^S(\mu, r, T^N) = 1 \text{ if } \mu \leq \underline{\mu}(r, T^N) = \sqrt{\underline{\mu}^2 (1 + T^N \frac{A}{r})}. \tag{67}$$

$$\delta^S(\mu, r, T^S) = \left(\frac{1}{\delta^S(\mu, r)} + T^S \frac{A}{r}\right)^{-1} \text{ and } \delta^N(\mu, r, T^S) = \delta^N(\mu, r)(1 + T^S \frac{A}{\mu^2}). \tag{68}$$

$$\delta^S(\mu, r, T^S) = 1 \text{ if } \mu \leq \underline{\mu}(r, T^S) = \sqrt{\underline{\mu}^2 (1 - T^S \frac{A}{r})} \tag{69}$$

$$\delta^N(\mu, r, T^S) = 1 \text{ if } \mu \geq \overline{\mu}(r, T^S) = \sqrt{\overline{\mu}^2 \left(1 - T^S \frac{A}{\overline{\mu}^2}\right)}. \tag{70}$$

$$\delta^N(\mu, r, T^N) \leq \delta^N(\mu, r) \text{ and } \delta^S(\mu, r, T^N) \geq \delta^S(\mu, r). \tag{71}$$

$$\overline{\mu}(r, T^N) \geq \overline{\mu} \text{ and } \underline{\mu}(r, T^N) \geq \underline{\mu}.$$

$$\delta^S(\mu, r, T^S) \leq \delta^S(\mu, r) \text{ and } \delta^N(\mu, r, T^S) \geq \delta^N(\mu, r). \tag{72}$$

$$\underline{\mu}(r, T^S) \leq \underline{\mu} \text{ and } \overline{\mu}(r, T^S) \leq \overline{\mu}. \tag{73}$$

$$\overline{T^N}(\mu, r, s_l, \delta) = \frac{s_l r(r-1)\left(2\mu^2\delta(2r-1)(4r-3) - r(4r-1)(3r-2)\right)}{2(4r-1)^2(3r-2)^2}. \tag{74}$$

$$\underline{T}^N(\mu, r, s_l, \delta) = \frac{s_l r(r-1)\left(\mu^2(4r-1)(3r-2) - 2\delta r(2r-1)(10r-7)\right)}{2\delta(4r-1)^2(3r-2)^2}. \tag{75}$$

$$\underline{T}^N(\mu, r, s_l, \delta) = -\frac{1}{\delta}\overline{T^S}(\mu, r, s_l, \delta). \tag{76}$$

$$\underline{T}^S(\mu, r, s_l, \delta) = -\frac{1}{\delta}\overline{T^N}(\mu, r, s_l, \delta). \tag{77}$$

$$\Delta T^i(r, s_l, \mu, \delta) \equiv \overline{T^i} - \underline{T}^i. \tag{78}$$

Autarky punishment

$$\delta^{Na}(\mu, r) = \frac{4\mu^2(r-1)(4r-1)}{\mu^2(4r-1)(13r-10) + 32r(3r-2)(r-1)}. \tag{79}$$

$$\delta^{Sa}(\mu, r) = \frac{4r^2(r-1)(4r-1)}{8\mu^2 r(3r-2)(r-1) - (4r-1)(4r-3)(4r^2-3r+2)}. \tag{80}$$

References

[1] Abreu, D., Pearce, D. and Stachetti, E. "Renegotiation and symmetry in repeated games." Journal of Economic Theory. 1993, 60, 217-40.

[2] Bagwell, K. and Staiger, R. W. “A Theory of Managed Trade.” American Economic Review, 1990, 80, 779-95.

[3] Dixit, A. "Strategic Aspects of Trade Policy", in T. F. Bewley ed., Advances in Economic Theory: Fifth World Congress, New York: Cambridge University Press, 1987.

[4] Farrell, J. and Maskin, E. "Renegotiation in repeated games." Games and Economic Behavior, 1989, 1, 327–360.

[5] Fudenberg D. and Tirole, J. Game Theory, MIT Press, Cambridge Massachusetts, 1991.

[6] Lahiri, S., Raimonod-Moller, P., Wong, K. and Woodland, A.D. "Optimal foreign aid and tariffs." Journal of Development Economics, 2002, 67, 79-99.

[7] Limao, N. M. and Saggi, K. "Tariff retaliation versus financial compensation in the enforcement of international trade agreements." Journal of International Economics, 2008, 76(1), 48-60.

[8] Ludema, R. D. “Optimal International Trade Agreements and Dispute-Settlement Procedures.” European Journal of Political Economy, 2001, vol 17(2), 355-376.

[9] Park J.-H. "International trade agreements between countries of asymmetric size." Journal of International Economics, 2000, 50, 473–495.

[10] Pearce, D. "Renegotiation-proof equilibria: collective rationality and intertemporal cooperation." Yale Cowles Foundation Discussion Paper: 855, 1987.

[11] Saggi, K. and Sara, N. "National Treatment at the WTO: the Roles of Product and Country Heterogeneity." *forthcoming* in International Economic Review, 2008.

[12] Shaked, A., and Sutton, J. “Relaxing Price Competition Through Product Differentiation.” Review of Economic Studies, 1982, 49, 3–13.