FREE TRADE NETWORKS WITH TRANSFERS*

By TAIJI FURUSAWA† and HIDEO KONISHI‡

†Hitotsubashi University ‡Boston College

We investigate the network of bilateral free trade agreements (FTAs) in the context of a network formation game with transfers. In a previous paper we showed that without international transfers countries with different industrialization levels cannot sign an FTA, so that the global free trade network, in which every pair of countries signs an FTA, is not in general pairwise-stable. In this paper we show that, even if the world consists of fairly asymmetric countries, the global free trade network is pairwise-stable when transfers between FTA signatories are allowed. Moreover, it is a unique pairwise-stable network unless industrial commodities are highly substitutable.
JEL Classification Numbers: C70, F15.

1. Introduction

Preferential trade agreements (PTAs) such as customs unions and free trade agreements (FTAs) have become increasingly popular among World Trade Organization (WTO) member countries. The WTO reports that over 170 PTAs are currently in force. It is easy to imagine that the web of PTAs worldwide is fairly complex.

The network formation game developed by Jackson and Wolinsky (1996) provides a suitable framework for the analysis of such a complex network of PTAs.[1] Goyal and Joshi (2001) and Furusawa and Konishi (2002) analyse FTA formation in the framework of the network formation game. The latter, for example, shows that if all countries are symmetric the global free trade network, in which every pair of countries forms an FTA, is pairwise-stable (Jackson and Wolinsky, 1996). Moreover, the global free trade network is the unique pairwise-stable network unless industrial commodities, which are subject to import tariffs when traded, are highly substitutable. Although they also analyse countries' incentives for FTAs in asymmetric countries, most of the strong results are obtained apply to symmetric countries.

Indeed, the asymmetry of countries is a major obstacle to FTA formation. Consider, for example, an FTA between a highly industrialized small country and a less industrialized large country. The FTA increases trade flows from the former to the latter disproportionately, dramatically increasing the trade surplus of the small, highly industrialized country and decreasing that of the large, less industrialized country. The direct impact of the FTA

---

[1] The network formation game is especially suitable for an analysis of FTAs, since it can deal with a hub-and-spoke system of FTA links. The hub-and-spoke system is a prominent feature of FTAs since, unlike customs unions, they allow member countries to impose different tariff rates on non-member countries.

on the industrial good trade surplus is favourable for the former and unfavourable for the latter. Unless opening the market benefits domestic consumers sufficiently, the less industrialized country will object to the FTA.

This rather intuitive discussion is formally supported with the aid of the welfare decomposition proposed by Furusawa and Konishi (2004), who show that, if each consumer's preferences are characterized by a quasi-linear utility function, social welfare can be decomposed into consumers' gross utility and the trade surplus of the non-numeraire good. Following Furusawa and Konishi (2002), the present paper adopts the quasi-linear utility function with a continuum of non-numeraire good commodities, which is developed by Ottaviano *et al.* (2002). Now the impact of an FTA on industrial good trade surplus can be decomposed into a direct surplus effect and a third-country effect. The former is the effect on the trade surplus of the partner country of the FTA in question, and the latter is the effect on the trade surplus of all other countries. The direct surplus effect can be positive or negative depending on the country's relative industrialization level compared with its partner country, as discussed above. On the other hand, the third-country effect is always positive since the country's exports to third countries are not affected by the FTA, while its imports from them decrease because such commodities become relatively more expensive in the domestic market.

Even if a pair of countries is so asymmetric that one of them has no incentive to sign an FTA, the FTA is likely to raise the joint welfare since the third-country effect is positive and the direct surplus effects are cancelled out between the two countries. (Each country's exports to the other are the other country's imports, and vice versa.) Therefore, if international transfers are possible they may sign an FTA, and the country that benefits from the FTA will transfer part of the benefit to the other country, which would incur a loss otherwise. International transfers are not uncommon in reality. A well known example is the monetary transfers that take place among EU members. The European Commission (2003) reports that in 2002 the "operational" budgetary balance of Germany, for example, was $-€5,067.8$ million, while that of Spain was $€8,870.8$ million, indicating that fairly large amounts of transfers are made among EU members. International transfers may not necessarily be made in the form of monetary transfers: recent FTAs have often included agreements on economic issues other than tariff reduction, such as competition policy, labour and environmental regulations. In many cases one signatory will demand that other signatories harmonize their standards with its own. Such unilateral changes of standards can be considered effectively as an international transfer.

We extend Furusawa and Konishi's (2002) model to allow international transfers between FTA signatories, in order to examine FTA formation between asymmetric countries. We find that their results are significantly strengthened if we allow international transfers. We obtained the same results for the global free trade network mentioned above, even where the world consists of fairly asymmetric countries. Mutual tariff elimination by any pair of countries benefits these countries as a whole unless industrial commodities are highly substitutable, so that they will sign an FTA if a transfer between them is possible. Hatta and Fukushima (1979) and Fukushima and Kim (1989) show that proportional reduction of all tariff rates of all countries improves world welfare.[2] Mutual tariff

---

[2]   Strictly speaking, world welfare improves if all the countries in the world (a) reduce all their *ad valorem* tariff rates proportionally (Hatta and Fukushima, 1979) or (b) reduce all their *specific* tariffs rates and subsidies proportionally (Fukushima and Kim, 1989). However, Kowalczyk (1989) demonstrates that world welfare may decline if all countries proportionally reduce all of their *ad valorem* tariff rates and subsidies.

reduction by an FTA is significantly different from their exercises, in that countries eliminate their tariffs only for their FTA partners. A discriminatory tariff profile caused by FTAs generally creates distortions in the domestic consumption of industrial commodities. If industrial commodities are not highly substitutable, these distortions are relatively small.

In this paper we propose two concepts that are suitable for network formation games with transfers. The first is a simple extension of Jackson and Wolinsky's (1996) pairwise stability. Roughly speaking, a pair of network and transfer systems is pairwise-stable in this context if there is a bilateral transfer system under which no country has an incentive to cut a link and no pair of unlinked countries have incentives to form a link with an appropriate transfer between themselves. The second concept is what we call the "contractual pairwise stability", whereby the effect of a link between a pair of countries spills over to other players. In our FTA network formation game, an FTA between a pair of countries adversely affects all other countries by reducing their exports to these two countries.[3] Then a country that has been an FTA partner of one of the new FTA signatories may be against this FTA, possibly threatening to terminate the existing link. We consider the situation in which countries need to compensate their FTA partners for possible resulting losses when they cut or form an FTA link. Roughly speaking, a pair of network and transfer systems is contractually pairwise-stable if it is pairwise-stable in this environment.

Two recent independent studies have also analysed transfers in network formation games, although, unlike ours, their approaches are based on non-cooperative games. Bloch and Jackson (2004) set up a general non-cooperative network formation game in which each player announces the amount of transfer (which can be negative) to each of other players. A pair of players form a link if and only if their proposed transfers to each other are compatible.[4] They first analyse the Nash equilibrium of this game, and then introduce its refinement, which they call the *pairwise equilibrium*. The set of outcomes in this equilibrium concept is a refinement of our cooperative game solution concept, i.e. pairwise stability with transfers. Matsubayashi and Yamakawa (2003) also analyse a non-cooperative link-cost sharing game in the context of Jackson and Wolinsky's (1996) connections model. They investigate whether there is an efficient Nash equilibrium in this game.

The rest of the paper is organized as follows. The next section sets out a multi-country model of the quasi-linear economy with a continuum of industrial commodities. Section 3 discusses incentives to sign a bilateral FTA and defines the pairwise stability in the presence of transfers. Section 4 derives main results about the global free trade network, and Section 5 shows that the same results obtain even with the contractual pairwise stability. Section 6 concludes.

2. The model

The basic model structure is the same as that of Furusawa and Konishi (2002), to which we refer for detailed derivation of the formulae presented below.

---

[3] Our FTA network formation game is what Goyal and Joshi (2003) call a "local spillover game". They define this as the game in which the marginal returns for the two players from the link depend only on the numbers of links that they each have.

[4] Bloch and Jackson (2004) also consider indirect transfer games, which allow player *i* to make transfers to players *j* and *k* contingent on the formation of a link between *j* and *k*.

2.1 Overview

Let $N = \{1, 2, \ldots, n\}$ be the set of n countries ($n \geq 2$), each of which is populated by a continuum of identical consumers who consume a numeraire good and a continuum of horizontally differentiated commodities that are indexed by $\omega \in [0, 1]$. A differentiated commodity can be considered as a variety of an industrial good. Each industrial commodity ω is produced by one firm which is also indexed by the same ω. We assume that there is no entry of firms to this industry. Each firm is owned equally by all domestic consumers who receive equal shares of all firms' profits. The numeraire good is produced competitively, on the other hand. Each consumer is endowed with l units of labour, which is used for the production of the industrial and numeraire goods. Each unit of labour produces one unit of the numeraire good, so that the wage rate equals 1. We also assume that industrial commodities are produced with a linear technology, and we normalize the unit labour requirement to be equal to 0 for each industrial commodity, without loss of generality. Alternatively, we can interpret the model such that each consumer is endowed with l units of the numeraire good, which can be transformed by a linear technology into industrial commodities.

In country $i \in N$, there are μ^i consumers and s^i firms, each producing a different industrial commodity. Thus, country i produces s^i different industrial commodities, which are consumed in every country in the world. We assume that the markets are segmented so that firms can perfectly price-discriminate among different countries. We normalize the size of total population so that $\sum_{k=1}^{n} \mu^k = 1$ as well as $\sum_{k=1}^{n} s^k = 1$. The ratio $\theta^i \equiv s^i/\mu^i$ measures country i's industrialization level: the higher the ratio, the higher the industrialization level. Country i imposes a specific tariff at a rate of t_k^i on the imports of industrial commodities produced in country k. The GATT Most Favoured Nation obligation will not allow any country to differentiate the tariff on like products based on originating countries except for products imported from partner countries of preferential trade agreements. Therefore we suppose that for country i the tariff rates are the same at t^i for all commodities imported from countries that have no FTA with country i. We assume that there is no commodity tax so that $t_i^i = 0$, and that the countries do not impose tariffs on the numeraire good, which may be traded internationally to balance the trade. Tariff revenue is redistributed equally to domestic consumers.

We examine the network of bilateral FTAs in the context of a network formation games with transfers.[5] An FTA between countries i and j is described by a *link*, which is an unordered pair of two countries. An *FTA graph* is an undirected graph, (N, Γ), consisting of the set of countries N and an FTA *network* Γ that is a collection of links. We say $\Gamma^{comp} = \{S \subset N : |S| = 2\}$ is a *complete network*, and the set of all networks is denoted by $\mathcal{G} = \{\Gamma \mid \Gamma \subseteq \Gamma^{comp}\}$. Let $C_i(\Gamma) = \{k \in N \mid t_k^i = 0 \text{ under } \Gamma\}$ represent the set of countries producing commodities on which country i does not impose tariffs. It follows from $t_i^i = 0$ that $C_i(\Gamma)$ is the set of country i's FTA partners and country i itself; i.e., $C_i(\Gamma) = \{i\} \cup \{k \in N : (i, k) \in \Gamma\}$. For simplicity, we omit the argument and write just C_i, as long as the network Γ is fixed.

---

[5] An FTA that involves more than two countries can be considered as a collection of bilateral FTAs between member countries.

2.2 Consumer demands and equilibrium

The utility function of a representative consumer in every country is given by the following quasi-linear utility function:

$$U(q, q_0) = \alpha \int_0^1 q(\omega)d\omega - \frac{\beta}{2}\int_0^1 q(\omega)^2 d\omega - \frac{\delta}{2}\left[\int_0^1 q(\omega)d\omega\right]^2 + q_0, \qquad (1)$$

where $q : [0, 1] \to \mathbb{R}_+$ is an integrable consumption function, and q_0 denotes the consumption level of the numeraire good.[6] The second-last term represents the substitutability among differentiated commodities, which may become clearer if we notice that $\left[\int_0^1 q(\omega)d\omega\right]^2 = \int_0^1\int_0^1 q(\omega)q(\omega')d\omega'd\cdot$ Letting y and $\tilde{p} : [0, 1] \to \mathbb{R}_+$ denote the consumer's income and the consumer price function, respectively, the budget constraint can be written as

$$y = \int_0^1 \tilde{p}(\omega)q(\omega)d\omega + q_0. \qquad (2)$$

Letting $p^i(\omega)$ and $\tilde{P}^i$ denote the producer price for commodity ω sold in country i and the average consumer price in country i, respectively, a representative consumer's demands in country i for a commodity ω produced in country k can be written as

$$q^i(\omega) = \frac{\alpha}{\beta} - \frac{1}{\beta}(p^i(\omega) + t_k^i) - \frac{\delta}{\beta(\beta + \delta)}(\alpha - \tilde{P}^i). \qquad (3)$$

Firm ω in country k chooses $\{p^i(\omega)\}_{i=1}^n$ in order to maximize its profits, $\pi(\omega) = \sum_{i=1}^n \mu^i p^i(\omega)q^i(\omega)$. The first-order condition for this maximization gives us

$$p^i(\omega) = \frac{1}{2}\left(\frac{\alpha\beta}{\beta + \delta} - t_k^i + \frac{\delta}{\beta + \delta}\tilde{P}^i\right) \qquad (4)$$

for any i. Notice that $p^i(\omega)$ does not vary with ω. Prices charged by firms depend only on the importing country's tariff policies. For simplicity, we henceforth suppress the argument ω.

It follows from (4) that country i's average consumer price $\tilde{P}^i$ is given by

$$\tilde{P}^i = \frac{\alpha\beta}{2\beta + \delta} + \frac{\beta + \delta}{2\beta + \delta}\bar{t}^i, \qquad (5)$$

where $\bar{t}^i \equiv \sum_{k=1}^n s^k t_k^i$. Substituting (5) into (4) yields the equilibrium producer price that each firm in country k charges for the market of country i, as a function of country i's tariff vector $\mathbf{t}^i = (t_1^i, \ldots, t_n^i)$:

$$p_k^i(\mathbf{t}^i) = \frac{\alpha\beta}{2\beta + \delta} - \frac{1}{2}t_k^i + \frac{\delta}{2(2\beta + \delta)}\bar{t}^i.$$

---

[6] This utility function is a continuous-goods version of those of Shubik (1984) and Yi (1996, 2000), who analyse the case where there are a finitely number of differentiated commodities. Our continuum-of-commodity setup is based on Ottaviano *et al.* (2002).

Then it follows from (3) that a representative consumer's demand in country i for a commodity produced in country k is

$$q_k^i(\mathbf{t}^i) = \frac{\alpha}{2\beta + \delta} - \frac{1}{2\beta}t_k^i + \frac{\delta}{2\beta(2\beta + \delta)}\bar{t}^i. \tag{6}$$

Notice that $p_k^i(\mathbf{t}^i) = \beta q_k^i(\mathbf{t}^i)$ holds for any tariff vector $\mathbf{t}^i$.

2.3 Social welfare

Under the world tariff vector $\mathbf{t} = (\mathbf{t}^1, \ldots, \mathbf{t}^n)$, each firm in country i earns profits

$$\pi_i(\mathbf{t}) = \sum_{k=1}^{n} \mu^k p_i^k(\mathbf{t}^k) q_i^k(\mathbf{t}^k) = \sum_{k=1}^{n} \mu^k \beta q_i^k(\mathbf{t}^k)^2. \tag{7}$$

Country i's *per capita* tariff revenue is

$$R^i(\mathbf{t}^i) = \sum_{k=1}^{n} t_k^i s^k q_k^i(\mathbf{t}^i). \tag{8}$$

A representative consumer's income in country i is the sum of labour income, redistributed tariff revenue and the profit shares of the firms in country i:

$$y = l + R^i(\mathbf{t}^i) + \frac{s^i \pi_i(\mathbf{t})}{\mu^i}. \tag{9}$$

Then it follows from (2) that

$$
\begin{aligned}
q_0^i(\mathbf{t}) &= l + R^i(\mathbf{t}^i) + \frac{s^i \pi_i(\mathbf{t})}{\mu^i} - \sum_{k=1}^{n} s^k[p_k^i(\mathbf{t}^i) + t_k^i]q_k^i(\mathbf{t}_k^i) \\
&= l + \sum_{k=1}^{n} s^k t_k^i q_k^i(\mathbf{t}^i) + \frac{s^i}{\mu^i}\sum_{k=1}^{n} \mu^k p_i^k(\mathbf{t}^k)q_i^k(\mathbf{t}^k) - \sum_{k=1}^{n} s^k[p_k^i(\mathbf{t}^i) + t_k^i]q_k^i(\mathbf{t}^i) \\
&= l - \sum_{k \neq i} s^k p_k^i(\mathbf{t}^i)q_k^i(\mathbf{t}^i) + \frac{s^i}{\mu^i}\sum_{k \neq i} \mu^k p_i^k(\mathbf{t}^k)q_i^k(\mathbf{t}^k),
\end{aligned}
$$

where $q^i(\omega) = q_k^i(\mathbf{t}^i)$ if ω is produced in country k.

Substituting this equilibrium demand into (1) and letting $q^i(\mathbf{t}^i)$ represent country i's equilibrium consumption function under the tariff $\mathbf{t}^i$, i.e. $q^i(\mathbf{t}^i) = (q_k^i(\mathbf{t}^i))_{k \in N}$, we obtain a representative consumer's utility in country i as a function of the world tariff vector. Country i's social welfare is the sum of the utilities across all consumers in country i.[7]

Lemma 1 (Furusawa and Konishi, 2002): *Country i's social welfare can be written as follows:*

$$W^i(\mathbf{t}) \equiv \mu^i U(q^i(\mathbf{t}^i), q_0^i(\mathbf{t})) = V^i(\mathbf{t}^i) + X^i(\mathbf{t}^{-i}) - M^i(\mathbf{t}^i), \tag{10}$$

---

[7] In the presence of transfers between countries, it is more convenient to work with (aggregate) social welfare than with per capita social welfare, as in Furusawa and Konishi (2002). The analysis itself is not affected by this modification.

where

$$V^i(\mathbf{t}^i) = \mu^i U(q^i(\mathbf{t}^i), l), \tag{11}$$

$$M^i(\mathbf{t}^i) = \mu^i \sum_{k \neq i} s^k p_k^i(\mathbf{t}^i) q_k^i(\mathbf{t}^i) = \beta \mu^i \sum_{k \neq i} s^k q_k^i(\mathbf{t}^i)^2, \tag{12}$$

$$X^i(\mathbf{t}^{-i}) = s^i \sum_{k \neq i} \mu^k p_i^k(\mathbf{t}^k) q_i^k(\mathbf{t}^k) = s^i \beta \sum_{k \neq i} \mu^k q_i^k(\mathbf{t}^k)^2, \tag{13}$$

with $\mathbf{t}^{-i} = (\mathbf{t}^1, \ldots, \mathbf{t}^{i-1}, \mathbf{t}^{i+1}, \ldots, \mathbf{t}^n)$. *The functions* $V^i(\mathbf{t}^i)$, $M^i(\mathbf{t}^i)$, *and* $X^i(\mathrm{t}^{-i})$ *represent consumers' gross utility, (industrial) import payments and (industrial) export profits, respectively.*[8] *Country i's social welfare consists of consumers' gross utility* $V^i(\mathbf{t}^i)$ *and the industrial trade surplus* $X^i(\mathbf{t}^{-i}) - M^i(\mathbf{t}^i)$.

Notice from (11)–(13) that an increase in a tariff rate affects $V^i(\mathbf{t}^i)$, $X^i(\mathbf{t}^{-i})$ and $M^i(\mathbf{t}^i)$ only through the changes in the consumption of industrial commodities. We see from (6) that the consumption of an industrial commodity depends on the tariff rate imposed on that commodity and the average tariff rate, i.e. $q_k^i(\mathbf{t}^i) \equiv \tilde{q}_k^i(t_k^i, \bar{t}^i)$. Thus we can write, for example, $V^i(\mathbf{t}^i) \equiv \tilde{V}^i(\tilde{q}_1^i(t_1^i, \bar{t}^i), \ldots, \tilde{q}_n^i(t_n^i, \bar{t}^i))$. An increase in t_j^i affects not only q_j^i directly, but also q_k^i indirectly, for all $k = 1, 2, \ldots, n$. These changes in consumption affect $V^i(\mathbf{t}^i)$ and $M^i(\mathbf{t}^i)$ in turn.

Country i's optimal tariff profile maximizes $V^i(\mathbf{t}^i) - M^i(\mathbf{t}^i)$ since $X^i(\mathbf{t}^{-i})$ does not depend on $\mathbf{t}^i$. We consider here the situation in which country i has signed FTAs, rather than CUs, with all other countries in C_i. Therefore, country i chooses its external tariffs without any coordination with other countries in C_i. As the following lemma shows, a country's optimal tariff rates depend only on its own characteristics, owing to the separability of a consumer's utility function.

Lemma 2 (Furusawa and Konishi, 2002): *Country i's optimal tariff rates are the same for all other countries. This common optimal tariff rate is a function of* s^i *and* $s^{C_i}(\equiv \sum_{k \in C_i} s^k)$:

$$t*(s^i, s^{C_i}) = \frac{4\alpha\beta(\beta + \delta s^i)}{3(2\beta + \delta)^2 - \delta(1 - s^{C_i})[4(2\beta + \delta) - \delta(1 - 2s^i)]} > 0,$$

which is increasing in s^i *and decreasing in* s^{C_i}.

3. Free trade agreements

3.1 Incentives to sign an FTA without transfers

We examine incentives for country i to sign an FTA with country j. If countries i and j sign an FTA, they eliminate all tariffs imposed on commodities imported from each other,

---

[8] See also Furusawa and Konishi (2004) for this welfare decomposition. Notice that the consumption of the numeraire good is fixed at l so that $V^i(\mathbf{t}^i)$ effectively measures the gross utility derived from the consumption of industrial commodities alone.

while keeping all other tariffs at their original levels. Letting $\mathbf{t}$ and $\mathbf{t}'$ denote the world tariff vectors before and after the FTA, respectively, $\mathbf{t}'$ differs from $\mathbf{t}$ only in that $t_j^{i\prime} = t_i^{j\prime} = 0$. When international transfers are not possible, country i has an incentive to sign an FTA with country j if and only if $W^i(\mathbf{t}') \geq W^i(\mathbf{t})$, which can be written as

$$\Delta V^i(\mathbf{t}^i) + [\Delta X^i(\mathbf{t}^{-i}) - \Delta M^i(\mathbf{t}^i)] \geq 0, \tag{14}$$

where Δ represents a change in the respective function values caused by an FTA between countries i and j such that $\Delta V^i(\mathbf{t}^i) \equiv V^i(\mathbf{t}^{i\prime}) - V^i(\mathbf{t}^i)$, for example. As we will see shortly, tariff reduction is likely to increase consumers' gross utility, unless the industrial commodities are highly substitutable. On the other hand, since the FTA increases the country's export profits and is also likely to increase the import payments, the FTA has an ambiguous impact on country i's industrial trade surplus.

Let us first investigate the sign of $\Delta V^i(\mathbf{t}^i)$. The next lemma shows that an FTA increases consumers' gross utility either if the substitutability among the industrial commodities is low, or if the original tariff rate is small.

Lemma 3 (Furusawa and Konishi, 2002): *A bilateral FTA with country j increases consumers' gross utility in country i, i.e. $\Delta V^i(\mathbf{t}^i) > 0$, if either of the following conditions is satisfied:*

$$(i) \quad 4\beta(\beta + \delta) - \delta^2(1 - 2s^{C_i} - s^j) \geq 0,$$

$$(ii) \quad t^i \leq \frac{8\alpha\beta^2}{\delta^2(1 - 2s^{C_i} - s^j) - 4\beta(\beta + \delta)}.$$

In particular, when country i levies a tariff rate t^i that is not greater than the optimal tariff rate $t^*(s^i, s^{C_i})$, it is sufficient that $\delta \leq 10\beta$ for $\Delta V^i(\mathbf{t}^i)$ to be positive.

These conditions reflect the second-best effect that partial removal of tax distortions may even reduce efficiency (Dixit, 1975; Hatta, 1977). In this case, the removal of country i's tariffs against country j enhances country i's consumers' gross utility if tariff distortions are not so widespread that $s^{C_i} + (1/2)s^j \geq 1/2$ holds (condition (i) in Lemma 3). Even when condition (i) is violated, the second-best effect is negligible if tax distortions are so small that condition (ii) is satisfied.

We next turn to the effect of an FTA between countries i and j on the industrial trade surplus. Let M_k^i and X_k^i be country i's import payments to country k and country i's export profits from country k, respectively:

$$M_k^i(\mathbf{t}^i) = \beta\mu^i s^k q_k^i(\mathbf{t}^i)^2,$$

$$X_k^i(\mathbf{t}^k) = \beta\mu^k s^i q_i^k(\mathbf{t}^k)^2 \ (= M_i^k(\mathbf{t}^k)).$$

Then we can rewrite country i's industrial trade surplus as

$$X^i(\mathbf{t}^{-i}) - M^i(\mathbf{t}^i) = \sum_{k \neq i} [X_k^i(\mathbf{t}^k) - M_k^i(\mathbf{t}^i)].$$

An FTA between i and j involves changes only in $\mathbf{t}^i$ and $\mathbf{t}^j$, so it does not affect $X_k^i(\mathbf{t}^k)$ for any $k \neq i, j$. Consequently, a change in country i's industrial trade surplus can be written as

$$\Delta[X^i(\mathbf{t}^{-i}) - M^i(\mathbf{t}^i)] = \underbrace{\Delta[X_j^i(\mathbf{t}^j) - M_j^i(\mathbf{t}^i)]}_{\text{direct surplus effect}} - \underbrace{\sum_{k \neq i,j} \Delta M_k^i(\mathbf{t}^i)}_{\text{third country effect}}.$$

The third-country effect, represented by the last term, is positive as long as country i's tariffs against countries other than j remain the same, since the reduction of t_j^i makes commodities imported from country j relatively less expensive, so that country i's imports from third countries decrease, i.e. $\Delta M_k^i(\mathbf{t}^i) < 0$. The third-country effect plays an important role in the later analysis of FTAs with transfers. It works positively for the FTA between countries i and j at the sacrifice of all other countries.

Having shown that the third-country effect is positive, let us now investigate the direct surplus effect, which can be rewritten as follows from the definitions of $M_i^j(\mathbf{t}^j)$ and $M_j^i(\mathbf{t}^i)$:

$$\underbrace{\Delta[X_j^i(\mathbf{t}^j) - M_j^i(\mathbf{t}^i)]}_{\text{direct surplus effect}} = \mu^i \mu^j \beta \Delta[\theta^i q_i^j(\mathbf{t}^j)^2 - \theta^j q_j^i(\mathbf{t}^i)^2],$$

where $\theta^i = s^i/\mu^i$ as defined above. The higher are θ^i and the lower θ^j, the larger will be an increase in country i's industrial trade surplus. That is, other things being equal, the relatively more industrialized country will be more enthusiastic than the less industrialized country about signing a bilateral FTA. The more industrialized country derives a larger benefit from the opening of the partner's relatively large market. In addition, opening its own market to the partner's firms does not significantly increase import payments, since the resulting penetration by the partner's firms is relatively small. Another important factor affecting incentives to sign an FTA is the difference in the original tariff rates. It is easy to see that if $t^i < t^j$, for example, then $\Delta q_j^i(\mathbf{t}^i) < \Delta q_i^j(\mathbf{t}^j)$. Country i's exports to country j increase more than its imports from country j, and hence the FTA between i and j tends to be more beneficial to country i.

The larger the difference in the two countries' characteristics, the greater is the direct surplus effect. Therefore, it is likely in such cases that the direct surplus effect for one of the countries takes a large negative value, outweighing the gains from the third-country effect and the possibly positive gross consumer utility effect. Then these countries will not sign an FTA (see discussions in Furusawa and Konishi, 2002). However, if transfers between them are allowed, the countries may overcome this problem. In the next subsection, we define our solution concept that is used to derive stable FTA networks in the presence of international transfers.

3.2 Stable free trade networks with transfers

If all tariffs are uniquely determined, exogenously or endogenously, for each free trade network $\Gamma \in \mathcal{G}$ (e.g. when all countries set their individual external tariffs at optimal levels, given the prevailing network Γ), then country i's social welfare without transfers is uniquely assigned to each Γ. We write this social welfare function as $u^i(\Gamma)$. We postpone determination of the function u^i until "tariff schemes" are discussed in the next section. The set of countries N and their payoff functions define a *network formation game*.

The following concept was first studied by Jackson and Wolinsky (1996). A *pairwise-stable network* is a network Γ^* such that (i) for any $i \in N$, and for any $(i, j) \in \Gamma^*$, $u^i(\Gamma^*)$

$\geq u^i(\Gamma^*\backslash(i, j))$; i.e., no country has an incentive to cut a link with another, and (ii) for any $(i, j) \notin \Gamma^*$ with $i \neq j$, if $u^i(\Gamma^*) < u^i(\Gamma^* \cup (i, j))$, then $u^j(\Gamma^*) > u^j(\Gamma^* \cup (i, j))$; i.e., for any pair of unlinked countries, at least one of them has no incentive to form a link with the other.

In this paper we examine FTA networks where international transfers between FTA signatories are allowed. This modification necessarily changes the definition of pairwise stability. Before we define pairwise stability in this context, let us first define transfers. Two countries that sign an FTA can make a transfer between them in order to compensate for any possible damage caused by the FTA. We require that transfers be made only within pairs of linked countries, so that transfers need to satisfy pairwise budget balancing condition.

Definition 1: *A **transfer from i to j** is the amount of the numeraire good $T_{ij} \in \mathbb{R}$ given from i to j such that $T_{ij} = -T_{ji}$. A **transfer system of the network** Γ is $T = (T_{ij})_{(i,j) \in \Gamma}$ such that $T_{ij} = -T_{ji}$ for any $(i, j) \in \Gamma$. A **network with transfers** (Γ, T) is a pair of a network and an associated transfer system.*[9]

Country i's payoff under (Γ, T) is given by $v^i(\Gamma, T) = u^i(\Gamma) + \sum_{j \in N} T_{ji}$. For a given (Γ, T) with $(i, j) \in \Gamma$, we let $(\Gamma \backslash (i, j), T_{-(ij,ji)})$ be a network of cutting (i, j) by cancelling transfers T'_{ij} and T'_{ji} between i and j without affecting any other links and transfers. For a given (Γ, T) with $(i, j) \notin \Gamma$ and $i \neq j$, we let $(\Gamma \cup (i, j), (T'_{ij}, T'_{ji}) \cup T)$ with $T'_{ij} = -T'_{ji}$ be a network of adding (i, j) with transfers T'_{ij} and T'_{ji} between i and j without affecting any other links and transfers.

Definition 2: *A network with transfers (Γ^*, T^*) is **pairwise-stable** if the following two conditions are satisfied simultaneously:*
 (i) For any $i \in N$, and for any $(i, j) \in \Gamma^*$, $v^i(\Gamma^*, T^*) \geq v^i(\Gamma^*\backslash(i, j), T^*_{-(ij,ji)})$; i.e., no country has an incentive to cut a link and thereby eliminate the transfer from (or to) that partner.
 (ii) For any $(i, j) \notin \Gamma^*$ with $i \neq j$, and for any (T'_{ij}, T'_{ji}) with $T'_{ij} = -T'_{ji}$, if $v^i(\Gamma^*, T^*) < v^i(\Gamma^* \cup (i, j), (T'_{ij}, T'_{ji}) \cup T^*)$, then $v^j(\Gamma^*, T^*) > v^j(\Gamma^* \cup (i, j), (T'_{ij}, T'_{ji}) \cup T^*)$; i.e., for any pair of unlinked countries, at least one of them has no incentive to form the link with any feasible transfer.

In concluding this section, we provide a useful characterization of pairwise-stable networks with transfers. The proof is relegated to the Appendix.

Lemma 4: *A network with transfers (Γ^*, T^*) is pairwise-stable if and only if the following two conditions are satisfied simultaneously:*
 (i) For any $(i, j) \in \Gamma^*$, $u^i(\Gamma^*) - u^i(\Gamma^*\backslash(i, j)) \geq T^*_{ij} \geq u^j(\Gamma^*\backslash(i, j)) - u^j(\Gamma^*)$.
 (ii) For any $(i, j) \notin \Gamma^*$ with $i \neq j$, $u^i(\Gamma^*) + u^j(\Gamma^*) \geq u^i(\Gamma^* \cup (i, j)) + u^j(\Gamma^* \cup (i, j))$.

Remark. For condition (i) in Lemma 4 to hold, it is necessary that $u^i(\Gamma^*) + u^j(\Gamma^*) \geq u^i(\Gamma^*\backslash(i, j)) + u^j(\Gamma^*\backslash(i, j))$ for any $(i, j) \in \Gamma^*$. Together with condition (ii), this observation

---

implies that, if transfers are allowed, then whether or not a pair of countries forms an FTA hinges on its impact on the joint social welfare of the two countries.[10]

4. Pairwise-stable free trade networks

As we have shown in the last section, the third-country effect is positive as long as tariff reforms are restricted to the mutual elimination of tariffs against each other. Moreover, we know that consumers' gross utility increases if either one of the conditions in Lemma 3 is satisfied. Therefore, what is left to be determined is the direct surplus effect.

Obviously, country i's imports from country j are country j's exports to country i, and hence we have $M_j^i(\mathbf{t}^i) = X_i^j(\mathbf{t}^i)$. Consequently,

$$\underbrace{\Delta[X_j^i(\mathbf{t}^j) - M_j^i(\mathbf{t}^i)]}_{i\text{'s direct surplus effect}} = -\underbrace{\Delta[X_i^j(\mathbf{t}^i) - M_i^j(\mathbf{t}^j)]}_{j\text{'s direct surplus effect}},$$

which implies that the sum of these two direct surplus effects is always zero. If transfers are allowed, it is sufficient for countries i and j to sign an FTA that ΔV^i and ΔV^j are positive (provided that all other tariffs than t_j^i and t_i^j remain the same).

To state our results formally, we define a tariff system that is consistent with FTA formation: A *tariff scheme* $(\tau_\Gamma^i)_{i\in N, \Gamma\in\mathcal{G}}$ is a list of functions $\tau_\Gamma^i : N \to \mathbb{R}$ such that $\tau_\Gamma^i(j) = t_j^i(\Gamma)$ for any $(i,j) \notin \Gamma$, and $\tau_\Gamma^i(j) = 0$ for any $(i,j) \in \Gamma$. Notice that τ_Γ^i shows country i's tariff profile under the FTA network Γ, and that a tariff scheme specifies every country's tariff profile under every possible FTA network.

With a fixed tariff scheme $(\tau_\Gamma^i)_{i\in N, \Gamma\in\mathcal{G}}$, we can formally define country i's (aggregate) social welfare under Γ, $u^i(\Gamma)$, for each $i \in N$ and $\Gamma \in \mathcal{G}$:

$$u^i(\Gamma) = V^i(\tau_\Gamma^i) + X^i((\tau_\Gamma^k)_{k\in N\setminus\{i\}}) - M^i(\tau_\Gamma^i). \tag{15}$$

To see the incentives for unlinked countries to form a new FTA, we define for $(i,j) \notin \Gamma$, $\Delta V^i \equiv V^i(\tau_{\Gamma\cup(i,j)}^i) - V^i(\tau_\Gamma^i)$, $\Delta X^i \equiv X^i((\tau_{\Gamma\cup(i,j)}^k)_{k\in N\setminus\{i\}}) - X^i((\tau_\Gamma^k)_{k\in N\setminus\{i\}})$, $\Delta M^i \equiv M^i(\tau_{\Gamma\cup(i,j)}^i) - M^i(\tau_\Gamma^i)$, and $\Delta u^i \equiv u^i(\Gamma \cup (i,j)) - u^i(\Gamma)$, and similarly for country j. Then, noticing that country i's exports to country j are country j's imports from country i, and vice versa, we have the following lemma. (The proof is relegated to the Appendix.)

Lemma 5: *For $(i,j) \notin \Gamma$, we have*

(i) $\displaystyle u^i(\Gamma) + u^j(\Gamma) = \left(V^i(\tau_\Gamma^i) + \sum_{k\neq i,j} X_k^i(\tau_\Gamma^k) - \sum_{k\neq i,j} M_k^i(\tau_\Gamma^i) \right)$

$\displaystyle + \left(V^j(\tau_\Gamma^j) + \sum_{k\neq i,j} X_k^j(\tau_\Gamma^k) - \sum_{k\neq i,j} M_k^j(T_\Gamma^j) \right),$

(ii) $\Delta u^i + \Delta u^j = \Delta V^i + \Delta V^j - \sum_{k\neq i,j} \Delta M_k^i - \sum_{k\neq i,j} \Delta M_k^j.$

---

[10] It may appear that our problem can be reformulated as a coalition formation game with partition functions. However, coalition formation games will not allow a country to be involved in multiple coalitions, which is a serious shortcoming in the analysis of FTA formation.

This lemma formally confirms the observation that whether a pair of countries benefits from forming an FTA depends only on the impacts on consumers' gross utilities and those on the imports from third countries.

Proposition 1: *For any arbitrary tariff scheme* $(\tau_\Gamma^i)_{i \in N, \Gamma \in \mathcal{G}}$, *there is a transfer system T associated with* Γ^{comp} *such that* (Γ^{comp}, T) *is a pairwise-stable network with transfers.*

Proof. Pick any pair $i, j \in N$ and let $\Delta u^i = u^i(\Gamma^{comp}) - u^i(\Gamma^{comp} \backslash (i, j))$ and similarly for j. We appeal to Lemma 5 to show that $\Delta u^i + \Delta u^j \geq 0$, and hence neither country has an incentive to cut the link. First, we show that the third-country effect is non-negative; i.e., $\sum_{k \neq i, j} \Delta M_k^i + \sum_{k \neq i, j} \Delta M_k^j \leq 0$. If $n = 2$, then there is no third-country effect. So let us now suppose that $n \geq 3$. For any $(\tau_\Gamma^i)_{i \in N, \Gamma \in \mathcal{G}}$, we know that all tariffs but t_j^i and t_i^j are zero when $\Gamma = \Gamma^{comp} \backslash (i, j)$. Thus, an FTA between i and j in $\Gamma^{comp} \backslash (i, j)$ will not change the third countries' tariff rates, which implies that $\sum_{k \neq i, j} \Delta M_k^i \leq 0$ and $\sum_{k \neq i, j} \Delta M_k^j \leq 0$. Next, we show that $\Delta V^i + \Delta V^j \geq 0$. Since C_i includes country i itself, we have $s^{C_i(\Gamma^{comp} \backslash (i, j))} + s^j = 1$, which immediately implies that $2s^{C_i(\Gamma^{comp} \backslash (i,j))} + s^j \geq 1$. Then condition (i) of Lemma 3 is satisfied so that $\Delta V^i > 0$. By the same argument, we also have $\Delta V^j > 0$. Hence we conclude that $\Delta u^i + \Delta u^j > 0$; i.e., $u^i(\Gamma^{comp}) - u^i(\Gamma^{comp} / (i, j)) \geq u^j(\Gamma^{comp} / (i, j)) - u^j(\Gamma^{comp})$. Then, there exists a transfer T_{ij} that satisfies condition (i) of Lemma 4.

Condition (ii) of Lemma 4 is vacuously satisfied since there is no unlinked pair under Γ^{comp}. ∎

Now, we seek a condition under which every pair of countries has its incentives for forming an FTA regardless of the existing FTA network. It is obvious that in such a case the complete network (global free trade) becomes a unique stable network.

Here we consider a specific tariff scheme, which we call the constant tariff scheme. The *constant tariff scheme* $(\tau_\Gamma^{ci})_{i \in N, \Gamma \in \mathcal{G}}$ is a tariff system by which any country specifies the same external tariff rate for any FTA network; i.e., for any $i \in N$, there is a constant t^i such that $\tau_\Gamma^{ci}(j) = t^i$ for any $j \in N$ with $(i, j) \notin \Gamma$. We now have the following proposition.

Proposition 2: *Suppose that* $\delta \leq 10\beta$; *i.e., the industrial commodities are not highly substitutable. Suppose also that the tariff system is the constant tariff scheme* $(\tau_\Gamma^{ci})_{i \in N, \Gamma \in \mathcal{G}}$ *such that each country's external tariff rate is no greater than its optimal tariff when there is no link; i.e.,* $\tau_\Gamma^{ci}(k) \leq t^*(s^i, s^i)$ *for any* $\Gamma \in \mathcal{G}$. *(Recall that* t^* *is decreasing in the second argument.) Then, under any FTA network with transfers* (Γ, T), *any unlinked pair of countries* i *and* j *has the incentive to sign an FTA with a transfer* T_{ij}' *between them. As a result, global free trade (the complete network* Γ^{comp}) *is a unique pairwise-stable network with transfers.*

Proof. For an arbitrary $\Gamma \in \mathcal{G} \backslash \Gamma^{comp}$, pick any $(i, j) \notin \Gamma$ and let $\Delta u^i = u^i(\Gamma \cup (i, j)) - u^i(\Gamma)$, and similarly for j. It follows from Lemma 3 that both ΔV^i and ΔV^j are positive. We also know that under the constant tariff scheme the third-country effects are non-negative; i.e., $\sum_{k \neq i, j} \Delta M_k^i \leq 0$ and $\sum_{k \neq i, j} \Delta M_k^j \leq 0$. Then we find from Lemma 5 that $\Delta u^i + \Delta u^j > 0$, or, equivalently, $u^i(\Gamma \cup (i, j)) + u^j(\Gamma \cup (i, j)) > u^i(\Gamma) + u^j(\Gamma)$. It follows from Lemma 4 that (Γ, T) is not pairwise-stable for any T. Indeed, there exists a transfer T_{ij}' such that $u^i(\Gamma \cup (i, j)) - u^i(\Gamma) \geq T_{ij}' \geq u^j(\Gamma) - u^j(\Gamma \cup (i, j))$, so that countries i and j have an incentive to sign an FTA with such a transfer T_{ij}'. ∎

The constant tariff scheme is a realistic tariff system although each country's tariff setting is necessarily suboptimal. What if countries optimally adjust their external tariffs when they sign a new FTA? Do we still have a result similar to Proposition 2?

To answer to this question, we consider a tariff system in which, when countries i and j sign an FTA, they select their individual external tariff rates t^i and t^j so as to maximize their joint social welfare, whereas all other countries keep their status quo tariffs.[11] In the presence of transfers, it is reasonable for countries that sign an new FTA to select their individual tariffs so as to maximize their joint welfare, since the transfer between them determines each country's share of the maximized "pie". Now, it follows from Lemma 5 that

$$\tilde{\mathbf{t}}^i \in \operatorname{argmax}\left[V^i(\mathbf{t}^i) - \sum_{k \neq i,j} M_k^i(\mathbf{t}^i) \right],$$

where $\tilde{\mathbf{t}}^i$ is country i's tariff profile such that $t_j^i = \tilde{t}^i$ if $j \notin C_i$ and $t_j^i = 0$ if $j \in C_i$; and similarly for j. It can be readily verified that $\tilde{t}^i = t^*(s^i + s^j, s^{C_i})$ and $\tilde{t}^j = t^*(s^i + s^j, s^{C_i})$.

If countries select their tariff rates in this manner, any pair of unlinked countries has an incentive to sign an FTA given that $\delta \leq 10\beta$. Proposition 2 implies that under any FTA network unlinked countries i and j have an incentive to sign an FTA if they do not adjust their tariffs. Now, under this tariff system countries have more incentives to sign an FTA, since they adjust their tariffs so as to maximize the joint social welfare. Therefore, the same result as Proposition 2 obtains even under this version of the optimal tariff system.

The significance of Proposition 2 and the following argument is that they apply regardless of countries' characteristics and existing FTA networks. As long as industrial commodities are not highly substitutable, any pair of unlinked countries always has an incentive to sign an FTA if they can make a transfer between themselves. No matter how different these countries' industrialization levels are, they have the incentive to sign an FTA. In this sense, FTAs are likely to be "building blocks" rather than "stumbling blocks" towards global free trade as long as transfers are allowed.[12]

We also find that FTA formation and multilateral trade negotiations such as GATT/ WTO trade negotiation rounds are complementary. The proof of Proposition 2 suggests that any pair of unlinked countries signs an FTA if their consumers' gross utilities increase as a consequence. However, Lemma 3 indicates that this is indeed the case if their current tariffs are low. Thus, multilateral trade negotiation may lower every involved country's tariff rates sufficiently for the continuous process of bilateral FTA formation to be kick-started. Moreover, Lemma 3 also indicates that bilateral FTA formation accelerates since condition (i) is more likely to be satisfied as more FTAs are formed.

---

[11] Country i's external tariff rate depends not only on Γ but also on the "path" to reach Γ, since it critically depends on the identity of the last country with which country i signed an FTA. Therefore, this tariff system is not a tariff scheme that we have formally defined.

[12] Although global free trade is not Pareto optimal in our oligopolistic model, it can be shown that a proportional reduction of all the tariff rates improves world welfare, as Fukushima and Kim (1989) show in a perfect competition model, so the movement towards global free trade is preferable as a whole. Since country i's import payments to country j is nothing but country j's export profits derived from country i, the world welfare equals $\sum_{i \in N} V^i(\mathbf{t}^i)$. It can be readily shown that, for any $i \in N$, $V^i(\mathbf{t}^i)$ increases as $\mathbf{t}^i$ proportionally decreases to 0.

5. Contractually stable free trade networks

We found in the last section that, unless industrial commodities are highly substitutable, the global FTA network is the unique pairwise-stable FTA. As Proposition 2 shows, any pair of countries has an incentive to sign an FTA, regardless of the current FTA network in such a situation. Then it appears that, if all countries are myopic, countries continue to sign bilateral trade agreements until global free trade is attained in an extended dynamic model. In the presence of international transfers, however, this may not be the case unless the transfer system continues to be adjusted as new FTA links are formed, in order to maintain countries' incentives to keep existing FTAs.

To see this argument, let us consider the situation in which every pair of countries has an incentive to sign an FTA as in Proposition 2. A new FTA link between i and j benefits these countries but adversely affects third countries, and thereby changes their incentives to form or cut FTA links. Consider country k, which has an FTA with country i in the FTA network Γ; so we have $u^k(\Gamma) - u^k(\Gamma \backslash (i, k)) \geq T_{ki}$. Country k has an incentive to keep FTA with country i even after the (i, j) link is formed if $u^k(\Gamma \cup (i, j)) - u^k((\Gamma \cup (i, j)) \backslash (i, k)) \geq T_{ki}$. Even though countries i and k jointly benefit from the (i, k) link, this condition may not hold unless T_{ki} is appropriately changed. If country k loses its incentive to keep the (i, k) link as a result of the (i, j) link, country i may hesitate to form a link with country j. Countries may not continue to form FTA links until global free trade is attained in the presence of international transfers.

This discussion suggests that, if countries compensate their FTA partners when they form new FTAs, they are likely to keep existing FTAs. If every pair of countries still benefits from a new FTA after full compensations for their partners' losses, we can expect the global FTA network to be finally attained in an extended dynamic game.

To examine this possibility, we introduce a new concept that captures the idea that, when a pair of countries signs an FTA, both must obtain "permission" from their respective FTA partners, possibly compensating for their resulting loss. In order to describe such compensations, we first define a new network with transfers obtained by cutting or forming a link from a network with transfers.

Definition 3:
(i) *A network with transfers* (Γ', T') *is* **agreeable at** (Γ, T) **with i's cutting of the link** (i, j) $\in \Gamma$*, if T' is the transfer system associated with $\Gamma' = \Gamma \backslash (i, j)$ such that (a) $T'_{ik} \geq T_{ik}$ (thus $T'_{ki} \leq T_{ki}$) if $k \in C_i(\Gamma) \backslash \{i, j\}$, and $T'_{lk} = T_{lk}$ for any other $(l, k) \in \Gamma$, and (b) $v^k(\Gamma, T) \leq v^k(\Gamma / (i, j), T')$ for any $k \in C_i(\Gamma) \backslash \{i, j\}$.*

(ii) *A network with transfers* (Γ', T') *is* **agreeable at** (Γ, T) **with a new link** $(i, j) \notin \Gamma$ *(for $i \neq j$), if T' is the transfer system associated with $\Gamma' = \Gamma \cup (i, j)$ such that (a) $T'_{ik} \geq T_{ik}$ (thus $T'_{ki} \leq T_{ki}$) if $k \in C_i(\Gamma) \backslash \{i\}$, $T'_{jk} \geq T_{jk}$ (thus $T'_{kj} \leq T_{kj}$) if $k \in C_j(\Gamma) \backslash \{j\}$, and $T'_{lk} = T_{lk}$ for any other $(l, k) \in \Gamma$, and (b) $v^k(\Gamma, T) \leq v^k(\Gamma \cup (i, j), T')$ for any $k \cup (C_i(\Gamma) \cup C_j(\Gamma)) \backslash \{i, j\}$.*

This definition requires that, when countries change the existing network with transfers either by unilaterally cutting a link or forming a link with another country, they can neither reduce their transfers to the original partners nor make their original partners worse off.

Definition 4: *A network with transfers* (Γ^*, T^*) *is* **contractually pairwise-stable** *if the following two conditions are simultaneously satisfied:*

(i) *For any $i \in N$, any $(i, j) \in \Gamma^*$, and any network with transfers (Γ', T') agreeable at (Γ^*, T^*) with i's cutting of the link (i, j), we have $v^i(\Gamma^*, T^*) \geq v^i(\Gamma', T')$.*

(ii) *For any $(i, j) \notin \Gamma^*$ with $i \neq j$, and for any network with transfers (Γ', T') agreeable at (Γ^*, T^*) with a new link (i, j), if $v^i(\Gamma^*, T^*) < v^i(\Gamma', T')$, then $v^j(\Gamma^*, T^*) > v^j(\Gamma', T')$.*

In the present environment, where the contractually pairwise stability is an appropriate solution concept, countries need to compensate their partners for possible losses caused by their decisions to form FTA links. On the other hand, even though their partners are better off by such decisions, they cannot change transfers to extract those gains.

Now, we have a counterpart of Lemma 4. The proof of the following lemma is similar to that of Lemma 4, and is relegated to the Appendix.

Lemma 6: *A network with transfers (Γ^*, T^*) is contractually pairwise-stable if and only if the following two conditions are simultaneously satisfied:*
(i) *For any $(i, j) \in \Gamma^*$,*

$$u^i(\Gamma^*) - u^i(\Gamma^* \backslash (i, j)) + \sum_{k \in C_i(\Gamma^*) \backslash \{i, j\}} \max\{u^k(\Gamma^*) - u^k(\Gamma^* \backslash (i, j)), 0\} \geq T^*_{ij}$$

$$\geq u^j(\Gamma^* \backslash (i, j)) - u^j(\Gamma^*) + \sum_{k \in C_j(\Gamma^*) / \{i, j\}} \min\{u^k(\Gamma^* / (i, j)) - u^k(\Gamma^*), 0\}.$$

(ii) *For any $(i, j) \notin \Gamma^*$ with $i \neq j$,*

$$\sum_{h = i, j} [u^h(\Gamma^* \cup (i, j)) - u^h(\Gamma^*)] \leq \sum_{k \in (C_i(\Gamma^*) \cup C_j(\Gamma^*)) \backslash \{i, j\}} \max\{u^k(\Gamma^*) - u^k(\Gamma^* \cup (i, j)), 0\}.$$

An FTA link (i, j) affects third countries' social welfare only through their exports to countries i and j. Since these exports (weakly) decline as a result of the (i, j) link, we have: (i) $u^k(\Gamma) - u^k(\Gamma \backslash (i, j)) \leq 0$ for any $k \in N \backslash \{i, j\}$ and for any Γ that includes (i, j); and (ii) $u^k(\Gamma) - u^k(\Gamma \cup (i, j)) \geq 0$ for any $k \in N \backslash \{i, j\}$ and for any Γ that does not include (i, j). Thus, we immediately obtain the following lemma.

Lemma 7: *In the FTA network formation game with transfers, (Γ^*, T^*) is contractually pairwise-stable if and only if the following two conditions are simultaneously satisfied*:
(i) *For any $(i, j) \in \Gamma^*$, $u^i(\Gamma^*) - u^i(\Gamma^* \backslash (i, j)) \geq T^*_{ij} \geq u^j(\Gamma^* \backslash (i, j)) - u^j(\Gamma^*)$.*
(ii) *For any $(i, j) \notin \Gamma^*$ and $i \neq j$, $\sum_{k \in C_i(\Gamma^*) \cup C_j(\Gamma^*)} [u^k(\Gamma^* \cup (i, j)) - u^k(\Gamma^*)] \leq 0$.*

Since cutting an FTA link benefits all other partner countries, a country need not compensate them when it eliminates an existing link. Therefore, condition (i) of Lemma 7 is the same as that of Lemma 4. However, a pair of countries must compensate their respective partner countries when they form a new FTA link. Their decision as to whether or not to sign an FTA depends not only on the impact on their own welfare, but also on their partners' welfare, as condition (ii) of Lemma 7 indicates.

Given that $u^k(\Gamma^* \cup (i, j)) - u^k(\Gamma^*) \leq 0$ for any $k \in (C_i(\Gamma^*) \cup C_j(\Gamma^*)) \backslash \{i, j\}$, condition (ii) of Lemma 7 is satisfied more easily than condition (ii) of Lemma 4. Since condition (i) is the same in both of these lemmas, we know that, if (Γ^*, T^*) is pairwise-stable, it is

also contractually pairwise-stable in the FTA network formation game with transfers. Consequently, we obtain the counterpart of Proposition 1.

Proposition 3: *For any arbitrary tariff scheme* $(\tau_\Gamma^i)_{i \in N, \Gamma \in \mathcal{G}}$*, there is a transfer system T associated with Γ^{comp} such that (Γ^{comp}, T) is a contractually pairwise-stable network with transfers.*

Somewhat surprisingly, the counterpart of Proposition 2 also obtains, since FTA signatories need only transfer part of gains from the third-country effect to compensate partner countries.

Proposition 4: *Suppose that $\delta \leq 10\beta$ and that the tariff system is the constant tariff scheme* $(\tau_\Gamma^{ci})_{i \in N, \Gamma \in \mathcal{G}}$*, such that each country's external tariff rate is not greater than its optimal tariff when there is no link. Then, under any FTA network with transfers (Γ, T), for any unlinked pair (i, j), there is an FTA network with transfers that is agreeable at (Γ, T) with a new link (i, j) and is preferred by both i and j to (Γ, T). Therefore, global free trade (the complete network Γ^{comp}) is a unique contractually pairwise-stable network with transfers.*

Proof. We show that, for any $\Gamma \in \mathcal{G} \backslash \Gamma^{comp}$, condition (ii) of Lemma 7 is violated if $\delta \leq 10\beta$. Let $\Delta V^i = V^i(\tau_{\Gamma \cup (i,j)}^i) - V^i(\tau_\Gamma^i)$ and similarly for other changes. Since the FTA between i and j only changes tariff schedules of these two countries, we have

$$\sum_{k \in C_i(\Gamma^*) \cup C_j(\Gamma^*)} [u^k(\Gamma^* \cup (i, j)) - u^k(\Gamma^*)]$$

$$= \Delta V^i + \Delta V^j - \sum_{k \neq i,j} (\Delta M_k^i + \Delta M_k^j) + \sum_{k \in (C_i(\Gamma^*) \cup C_j(\Gamma^*)) \backslash \{i,j\}} (\Delta X_i^k + \Delta X_j^k)$$

$$= \Delta V^i + \Delta V^j - \sum_{k \notin C_i(\Gamma^*) \cup C_j(\Gamma^*)} (\Delta M_k^i + \Delta M_k^j),$$

$$> 0,$$

where we have used Lemma 3 and the fact that the (partial) third-country effect is positive. ∎

6. Concluding remarks

We have analysed bilateral FTA networks of the world by a network formation game with transfers. When transfers are allowed between FTA signatories, a pair of countries forms an FTA if and only if their *joint* welfare improves as a consequence, so that countries are more likely to sign an FTA. Indeed, we find that the complete FTA network, which effectively attains the global free trade, is pairwise-stable. Surprisingly, this result obtains even if countries are far from symmetric, sharply contrasting with Furusawa and Konishi (2002), who derive a similar result in the case of symmetric countries when transfers are not possible. Moreover, if industrial commodities are not highly substitutable, the global free trade network is the unique pairwise-stable network. In that case every pair of countries has an incentive to sign an FTA.

Since the direct surplus effects are cancelled out between the FTA signatories, and their (industrial) trade surplus with third countries improve as a result of the substitution effect

of industrial commodities, a bilateral FTA enhances the joint welfare of the signatories if consumers' gross utilities from industrial commodities jointly increase. This situation is likely to arise if the FTA does not increase distortion in the consumption of industrial commodities (the second-best effect). The distortion would be small if industrial commodities were not highly substitutable. The resulting distortion can also be negligible if countries' external tariffs are low thanks to the past multilateral trade negotiation under the auspices of the GATT/WTO. As Freund (2000) demonstrates, multilateral trade negotiations and regional trade liberalizations based on the GATT Article XXIV are complementary in this sense even though they appear contradictory when viewed from the perspective of the GATT principle of non-discrimination.

When a pair of countries signs an FTA, all other countries including their FTA partners are worse off. Therefore a country may well be against its partner-country's new FTA, possibly threatening to terminate the existing FTA between them. When transfers are possible, countries may compensate partner-countries for such possible losses when they form new FTAs. We have proposed contractually pairwise stability as a suitable solution to this problem. We have shown somewhat surprisingly that the results about the pairwise-stable global free trade network obtain even if we use the contractually pairwise stability as the solution concept.

In concluding, we want to point out that we can derive qualitatively the same result in an extended model that includes more than one industrial good. Of course, including additional goods adds more dimensions of similarity between countries and hence enables us to analyse more problems of interest. For the brevity of the current analysis, however, we leave this extension for future research.

Appendix

A1 Proof of Lemma 4

First, we show that the definition of pairwise-stable networks with transfers implies the conditions in this lemma. We start with condition (i). Let $(i, j) \in \Gamma^*$. Then, we have $v^i(\Gamma^*, T^*) \geq v^i(\Gamma^* \backslash (i, j), T^*_{-(ij,ji)})$ and $v^j(\Gamma^*, T^*) \geq v^j(\Gamma^* \backslash (i, j), T^*_{-(ij,ji)})$ from Definition 2. Rewriting the former inequality, we have

$$u^i(\Gamma^*) + \sum_{k \in C_i(\Gamma^*) \backslash \{i\}} T^*_{ki} \geq u^i(\Gamma^* \backslash (i, j)) + \sum_{k \in C_i(\Gamma^*) \backslash \{i, j\}} T^*_{ki}$$

$$T^*_{ji} \geq u^i(\Gamma^* \backslash (i, j)) - u^i(\Gamma^*).$$

Similarly, the latter inequality with regard to country j is equivalent to

$$T^*_{ij} \geq u^j(\Gamma^* \backslash (i, j)) - u^j(\Gamma^*).$$

Then condition (i) follows from $T^*_{ij} = -T^*_{ji}$.

Next, we show that condition (ii) holds for any pairwise-stable network with transfers. Suppose to the contrary that $u^i(\Gamma^*) + u^j(\Gamma^*) < u^i(\Gamma^* \cup (i, j)) + u^j(\Gamma^* \cup (i, j))$. Let $\varepsilon \equiv u^i(\Gamma^* \cup (i, j)) + u^j(\Gamma^* \cup (i, j)) - u^i(\Gamma^*) - u^j(\Gamma^*) > 0$, and let $T'_{ji} = u^i(\Gamma^*) - u^i(\Gamma^* \cup (i, j)) + \varepsilon/2$. Then we have from $T'_{ij} + T'_{ji} = 0$ that $T'_{ij} = u^j(\Gamma^*) - u^j(\Gamma^* \cup (i, j)) + \varepsilon/2$. Now, using $T^*_{ji} = 0$, we have

$$v^i(\Gamma^* \cup (i, j), (T'_{ij}, T'_{ji}) \cup T^*) = u^i(\Gamma^* \cup (i, j)) + \sum_{k \in N \setminus \{j\}} T^*_{ki} + T'_{ji}$$

$$= u^i(\Gamma^*) + \sum_{k \in N} T^*_{ki} + \frac{\varepsilon}{2} \qquad (A1)$$

$$= v^i(\Gamma^*, T^*) + \frac{\varepsilon}{2}.$$

Similarly, we have

$$v^j(\Gamma^* \cup (i, j), (T'_{ij}, T'_{ji}) \cup T^*) = v^j(\Gamma^*, T^*) + \varepsilon/2. \qquad (A2)$$

Equalities (A1) and (A2) imply that countries i and j both gain from forming a new link with the transfer system $(T'_{ij}, T'_{ji}) \cup T^*$, which contradicts the supposition that (Γ^*, T^*) is pairwise-stable.

Now, we show that conditions (i) and (ii) of this lemma imply that (Γ^*, T^*) is pairwise-stable. We start by demonstrating that condition (i) implies condition (i) of Definition 2. Let $(i, j) \in \Gamma^*$. Then, country i's social welfare with transfers satisfies the following:

$$v^i(\Gamma^*, T^*) = u^i(\Gamma^*) + \sum_{k \in N \setminus \{j\}} T^*_{ki} + T^*_{ji}$$

$$= u^i(\Gamma^*) - u^i(\Gamma^* \setminus (i, j)) + u^i(\Gamma^* \setminus (i, j)) + \sum_{k \in N \setminus \{j\}} T^*_{ki} + T^*_{ji}$$

$$= u^i(\Gamma^*) - u^i(\Gamma^* \setminus (i, j)) + T^*_{ji} + v^i(\Gamma^* \setminus (i, j), T^*_{-(ij, ji)})$$

$$\geq v^i(\Gamma^* \setminus (i, j), T^*_{-(ij, ji)}),$$

where the last inequality follows from condition (i) of the lemma.

Let us turn to condition (ii). Let $(i, j) \notin \Gamma^*$ with $i \neq j$, and select an arbitrary pair of transfers (T'_{ij}, T'_{ji}) with $T'_{ij} = -T'_{ji}$ such that $v^i(\Gamma^*, T^*) < v^i(\Gamma^* \cup (i, j), (T'_{ij}, T'_{ji}) \cup T^*)$. This inequality is equivalent to $u^i(\Gamma^*) + \sum_{k \in N} T^*_{ki} < u^i(\Gamma^* \cup (i, j)) + \sum_{k \in N \setminus \{j\}} T^*_{ki} + T'_{ji}$. Then it follows from $T^*_{ji} = 0$ and $T'_{ij} = -T'_{ji}$ that $T'_{ij} < u^i(\Gamma^* \cup (i, j)) - u^i(\Gamma^*)$. Consequently, we obtain from condition (ii) that $u^j(\Gamma^*) - u^j(\Gamma^* \cup (i, j)) \geq u^i(\Gamma^* \cup (i, j)) - u^i(\Gamma^*) > T'_{ij}$. Then, we immediately obtain $v^j(\Gamma^*, T^*) > v^j(\Gamma^* \cup (i, j), (T''_{ij}, T'_{ji}) \cup T^*)$, which shows that condition (ii) is satisfied. ∎

A2 Proof of Lemma 5

Since country i's exports to country j are country j's imports from country i, and vice versa, we have $X^i_j = M^j_i$ and $X^j_i = M^i_j$. Statement (i) follows immediately. Moreover, these countries' exports to the third countries do not change as a result of the FTA between i and j; i.e., $\Delta X^i_k = \Delta X^j_k = 0$ for any $k \in N \setminus \{i, j\}$. Consequently, we obtain from (i) that

$$\Delta u^i + \Delta u^j = \Delta V^i + \Delta V^j + \Delta(X^i_j - M^i_j) + \sum_{k \neq i, j} \Delta(X^i_k - M^i_k)$$

$$+ \Delta(X^j_i - M^j_i) + \sum_{k \neq i, j} \Delta(X^j_k - M^j_k)$$

$$= \Delta V^i + \Delta V^j - \sum_{k \neq i, j} \Delta M^i_k - \sum_{k \neq i, j} \Delta M^j_k. \ ∎$$

A3   Proof of Lemma 6

First, we show that Definition 4 (i) implies Lemma 6 (i) for country i. We rewrite the condition $v^k(\Gamma^*, T^*) \leq v^k(\Gamma^* \backslash (i, j), T')$ for $k \in C_i(\Gamma^*) \backslash \{i, j\}$ as

$$u^k(\Gamma^*) + \sum_{l \in N} T_{lk}^* \leq u^k(\Gamma^* \backslash (i, j)) + \sum_{l \in N} T_{lk}'$$

$$u^k(\Gamma^*) + T_{ik}^* \leq u^k(\Gamma^* \backslash (i, j)) + T_{ik}'$$

$$T_{ik}' \geq T_{ik}^* + u^k(\Gamma^*) - u^k(\Gamma^* \backslash (i, j)),$$

where we have used $T_{lk}' = T_{lk}^*$ for $l \neq i$. Then it follows from $T_{ik}' \geq T_{ik}^*$ that

$$T_{ik}' \geq T_{ik}^* + \max\{u^k(\Gamma^*) - u^k(\Gamma^*/(i, j)), 0\}. \tag{A3}$$

We fix T_{ik}' at the value that satisfies (A3) with equality. Then we can rewrite $v^i(\Gamma^*, T^*) \geq v^i(\Gamma^* \backslash (i, j), T')$ as

$$u^i(\Gamma^*) - T_{ij}^* - \sum_{k \in C_i(\Gamma^*) \backslash \{i, j\}} T_{ik}^* \geq u^i(\Gamma^* \backslash (i, j)) - \sum_{k \in C_i(\Gamma^*) \backslash \{i, j\}} [T_{ik}^* + \max\{u^k(\Gamma^*) - u^k(\Gamma^* \backslash (i, j)), 0\}],$$

which gives us the first part of Lemma 6 (i). The second part regarding country j obtains similarly.

Next, we show that Lemma 6 (i) implies Definition 4 (i). Tracing the calculation in the above backward, we obtain

$$u^i(\Gamma^*) - \sum_{k \in C_i(\Gamma^*) \backslash \{i\}} T_{ik}^* \geq u^i(\Gamma^* \backslash (i, j)) - \sum_{k \in C_i(\Gamma^*) \backslash \{i, j\}} [T_{ik}^* + \max\{u^k(\Gamma^*) - u^i(\Gamma^* \backslash (i, j)), 0\}].$$

It then follows from (A3) that

$$u^i(\Gamma^*) - \sum_{k \in C_i(\Gamma^*) \backslash \{i\}} T_{ik}^* \geq u^i(\Gamma^* \backslash (i, j)) - \sum_{k \in C_i(\Gamma^*) \backslash \{i, j\}} T_{ik}'$$

$$v^i(\Gamma^*, T^*) \geq v^i(\Gamma^* \backslash (i, j), T').$$

A similar result obtains for country j.

Now, we demonstrate that Definition 4 (ii) implies Lemma 6 (ii) by showing the contrapositive. Suppose to the contrary that

$$\sum_{h = i, j} [u^h(\Gamma^* \cup (i, j)) - u^h(\Gamma)] > \sum_{k \in (C_i(\Gamma^*) \cup C_j(\Gamma^*)) \backslash \{i, j\}} \max\{u^k(\Gamma^*) - u^k(\Gamma^* \cup (i, j)), 0\}.$$

For any $k \in (C_i(\Gamma^*) \cup C_j(\Gamma^*)) \backslash \{i, j\}$, we have from $v^k(\Gamma^*, T^*) \leq v^k(\Gamma^* \cup (i, j), T')$ that

$$T_{ik}' + T_{jk}' \geq T_{ik}^* + T_{jk}^* + u^k(\Gamma^*) - u^k(\Gamma^* \cup (i, j)).$$

Then it follows from $T_{ik}' \geq T_{ik}^*$ and $T_{jk}' \geq T_{jk}^*$ that

$$T_{ik}' + T_{jk}' \geq T_{ik}^* + T_{jk}^* + \max\{u^k(\Gamma^*) - u^k(\Gamma^* \cup (i, j)), 0\}. \tag{A4}$$

Let $T'_{ik} + T'_{jk}$ be the value that satisfies (A4) with equality. Then we have

$$\sum_{h=i,j} [u^h(\Gamma^* \cup (i, j)) - u^h(\Gamma^*)] > \sum_{(C_i(\Gamma^*) \cup C_j(\Gamma^*)) \backslash \{i,j\}} (T'_{ik} + T'_{jk} - T^*_{ik} - T^*_{jk}),$$

which gives us

$$\left[u^i(\Gamma^* \cup (i, j)) - \sum_{k \in C_i(\Gamma^*)} T'_{ik} \right] - \left[u^i(\Gamma^*) - \sum_{k \in C_i(\Gamma^*)} T^*_{ik} \right]$$

$$+ \left[u^j(\Gamma^* \cup (i, j)) - \sum_{k \in C_j(\Gamma^*)} T'_{jk} \right] - \left[u^j(\Gamma^*) - \sum_{k \in C_j(\Gamma^*)} T^*_{jk} \right] > 0.$$

Let ε be the left-hand side of the above inequality, and let T'_{ij} be such that

$$T'_{ij} = \left[u^i(\Gamma^* \cup (i, j)) - \sum_{k \in C_i(\Gamma^*)} T'_{ik} \right] - \left[u^i(\Gamma^*) - \sum_{k \in C_i(\Gamma^*)} T^*_{ik} \right] - \frac{\varepsilon}{2}.$$

Then we have $v^i(\Gamma^* \cup (i, j), T') = v^i(\Gamma^*, T^*) + \varepsilon/2$. Moreover, we have from the definition of ε that $v^j(\Gamma^* \cup (i, j), T') = v^j(\Gamma^*, T^*) + \varepsilon/2$, which violates Definition 4 (ii).

Finally, we show that Lemma 6 (ii) implies Definition 4 (ii). Again, we show the contrapositive of the statement. Suppose there exists a transfer system T' such that $v^i(\Gamma^* \cup (i, j), T') > v^i(\Gamma^*, T^*)$ and $v^j(\Gamma^* \cup (i, j), T') \geq v^j(\Gamma^*, T^*)$, satisfying all other requirements. Then we obtain

$$v^i(\Gamma^* \cup (i, j), T') + v^j(\Gamma^* \cup (i, j), T') > v^i(\Gamma^*, T^*) + v^j(\Gamma^*, T^*),$$

which is reduced to

$$\sum_{h=i,j} [u^h(\Gamma^* \cup (i, j)) - u^h(\Gamma^*)] > \sum_{k \in C_i(\Gamma^*) \cup C_j(\Gamma^*)} (T'_{ik} + T'_{jk} - T^*_{ik} - T^*_{jk}).$$

Then it follows from (A4) that

$$\sum_{h=i,j} [u^h(\Gamma^* \cup (i, j)) - u^h(\Gamma^*)] > \sum_{k \in C_i(\Gamma^*) \cup C_j(\Gamma^*)} \max\{u^k(\Gamma^*) - u^k(\Gamma^* \cup (i, j)), 0\}.$$

REFERENCES

Bloch, F. and M. O. Jackson (2004) *The Formation of Networks with Transfers among Players*, Working Paper, California Institute of Technology.

Dixit, A. K. (1975) "Welfare Effect and Tax and Price Changes", *Journal of Public Economics*, Vol. 4, pp. 103–123.

European Commission (2003) "Allocation of 2002 EU Operating Expenditure by Member State, at http://europa.eu.int/comm/budget/agenda2000/reports_en.htm.

Freund, C. L. (2000) "Multilateralism and the Endogenous Formation of Preferential Trade Agreements", *Journal of International Economics*, Vol. 52, pp. 359–376.

Fukushima, T. and N. Kim (1989) "Welfare Improving Tariff Changes: A Case of Many Goods and Countries", *Journal of International Economics*, Vol. 26, pp. 383–388.

Furusawa, T. and H. Konishi (2002) *Free Trade Networks*, Working Paper No. 548, Boston College, Mass.

—— and —— (2004) "A Welfare Decomposition in Quasi-Linear Economies", *Economics Letters*, Vol. 85, pp. 29–34.

Goyal, S. and S. Joshi (2001) "Bilateralism and Free Trade", unpublished paper.

—— and —— (2003) "Unequal Connections", unpublished paper.

Hatta, T. (1977) "A Theory of Piecemeal Policy Recommendations", *Review of Economic Studies*, Vol. 44, pp. 1–21.

—— and T. Fukushima (1979) "The Welfare Effect of Tariff Rate Reductions in a Many-Country World", *Journal of International Economics*, Vol. 9, pp. 503–511.

Jackson, M. O. and A. Wolinsky (1996) "A Strategic Model of Social and Economic Networks", *Journal of Economic Theory*, Vol. 71, pp. 44–74.

Kowalczyk, C. (1989) "Trade Negotiations and World Welfare", *American Economic Review*, Vol. 79, pp. 552–559.

Matsubayashi, N. and S. Yamakawa (2003) *A Network Formation Game with an Endogenous Cost Allocation Rule*, Working Paper, Tokyo University of Science.

Ottaviano, G. I. P., T. Tabuchi and J.-F. Thisse (2002) "Agglomeration and Trade, Revisited", *International Economic Review*, Vol. 43, pp. 409–436.

Shubik, M. (1984) *A Game-Theoretic Approach to Political Economy*, Cambridge, Mass.: MIT Press.

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

—— (2000) "Free-Trade Areas and Welfare: An Equilibrium Analysis", *Review of International Economics*, Vol. 8, pp. 336–347.