Farsightedly Stable FTA Structures

Nakanishi, Noritsugu

Farsightedly Stable FTA Structures[*]

Noritsugu NAKANISHI[†]

Graduate School of Economics, Kobe University
Rokkodai-cho 2-1, Nada-ku, Kobe 657-8501
JAPAN

This version compiled: July 20, 2011[‡]

Abstract

Taking account of the farsightedness of the countries and adopting the von Neumann-Morgenstern (vNM) stable set as the solution concept, we examine an FTA network formation game. FTA networks are represented by undirected graphs with their vertex sets being identified with the set of the countries. Each country's welfare depends upon the shape of the graph and its location in that graph. We examine two extreme cases: one in which the pre-agreement tariffs are very high and the other in which they are sufficiently low. In the former, the farsighted vNM stable set only supports global free trade, implying that bilateral FTAs are *building blocks* for achieving global free trade. In the latter, the farsighted vNM stable set does not support global free trade, instead it supports some inefficient FTA networks, implying that bilateral FTAs are *stumbling blocks* against achieving global free trade.

[*]An earlier version of this paper ("Stable FTA Structure—An Application of the von Neumann-Morgenstern Stable Set to International Trade") was presented at the 6th Pan-Pacific Conference on Game Theory, Tokyo Institute of Technology (February 28–March 2, 2011). The author is grateful to Professors Eric Bond, Ngo Van Long, Yasukazu Ichino, and Toru Kikuchi for their valuable comments and suggestions. He also acknowledges the financial support of the Japan Society for the Promotion of Science [Grant-in-Aid for Scientific Research (A), No. 22243024].

[†]Address for correspondence: Noritsugu Nakanishi, Graduate School of Economics, Kobe University, Rokkodai-cho 2-1, Nada-ku, Kobe 657-8501, JAPAN. Tel: +78-803-6837 (Office Direct). Fax: +78-803-7293 (Department Office). E-mail: `nakanishi@econ.kobe-u.ac.jp`

[‡]The first version: October 10, 2010; the second version: February 4, 2011.

1 Introduction

Since the late 1980s, preferential (or regional) trade agreements such as customs unions (CUs) and free trade agreements (FTAs) have been growing rapidly in number and becoming prevalent move in the international trade scene. Well-known examples include the European Union (EU), the North American Free Trade Agreement (NAFTA), the South American Common Markets (MERCOSUR), and the Association of South-East Asian Nations (ASEAN). Even Japan, who had long been advocating for a multilateral approach to trade liberalization under the GATT/WTO regime, has recently concluded some bilateral FTAs, which are sometimes referred to as the Economic Partnership Agreements (EPAs), with Singapore, Mexico, Malaysia, Philippine, Chile, and other countries.[1]

Bhagwati (1991, 1993) has argued that this trend of regionalism/bilateralism is harmful to the accomplishment of *global free trade* under the auspices of the GATT/WTO regime, which has been traditionally considered to be *efficient* and/or *welfare-enhancing* from the point of view of the world as a whole. Further, he has raised questions as to whether preferential trading blocs (CUs and/or FTAs) reduce or increase the world welfare and whether the prevalence of bilateralism/regionalism can eventually lead to the situation where in the world welfare is maximized. In other words, he raises a question as to whether preferential trading blocs can be a "building block" for or a "stumbling block" against the achieving of global free trade. The former can be traced back to Viner (1950)'s question that preferential trading blocs can be trade-creating or trade-diverting. The latter is now known as the "dynamic time-path" question.[2]

To answer these questions raised by Bhagwati (1991, 1993), a considerable number of theoretical as well as empirical literature has emerged. Recent theoretical studies (on the "dynamic time-path" question) can be split into two interesting approaches: one is based on the coalition/network formation games and the other

[1]WTO has reported in its website (as of October 15, 2010) that "The surge in RTAs has continued unabated since the early 1990s. As of 31 July 2010, some 474 RTAs, counting goods and services notifications separately, have been notified to the GATT/WTO. Of these, 351 RTAs were notified under Article XXIV of the GATT 1947 or GATT 1994; 31 under the Enabling Clause; and 92 under Article V of the GATS. At that same date, 283 agreements were in force."

[2]In this paper, we mainly focus on the latter dynamic time-path question.

is based on some "dynamic" games.[3]

For example, Yi (1996, 2000), Das and Ghosh (2006), and Saggi and Yildiz (2010) have followed the line of coalition formation game approach. Yi (1996) has constructed a model with *ex ante* symmetric countries, in which countries intend to form CUs of certain sizes. In his model, global free trade is *efficient* (in the sense that the world welfare is maximized), formation of CUs exhibits *negative externalities* on the welfare of the outsider countries, and international income transfers are not allowed. He has shown that global free trade is the unique Nash equilibrium outcome of the simultaneous move, open regionalism game. He has also shown that although global free trade is the only *symmetric* outcome of the subgame perfect Nash equilibrium (SPNE) of the sequential move, open regionalism game, a typical SPNE outcome is *asymmetric* such that it contains the largest customs union that is unique as well as some other smaller CUs.[4] Das and Ghosh (2006) have considered a four-country model of coalition formation. They have assumed asymmetry among countries in terms of income levels generated by the difference in human capital content: two of them are high-income North countries and the other two are low-income South countries. Adopting the coalition-proof Nash equilibrium (CPNE) as the solution concept, they have shown that if the market size difference (the difference in income levels) is sufficiently high, only the North-North pair and the South-South pair form FTAs (i.e., polarization), but no North-South pair is realized in the CPNE; otherwise, global free trade is realized in the CPNE. Saggi and Yildiz (2010) have constructed a three-country model, and investigated both bilateralism and multilateralism games. Assuming *ex ante* symmetry among countries, they have shown that global free trade is the only CPNE outcome in both games.

On the other hand, Furusawa and Konishi (2005, 2007) and Goyal and Joshi (2006) have formulated the situation within the framework of network formation games as developed by Jackson and Wolinsky (1996). They have examined the in-

[3]Because there is a vast amount of literature concerning the dynamic time-path question, we only consider those studies that are the most relevant to the current paper.

[4]Yi (2000) has considered a similar model in which regional trade agreements take the form of FTAs instead of CUs. There is an important difference between the cases of CUs and FTAs in his model: formation of CUs exhibits "negative" externalities on the welfare of the outsider countries, while that of FTAs exhibits "positive" externalities.

centives of countries to form or dissolve bilateral FTAs. Then, adopting the notion of *pairwise stability* as the solution concept, they have shown (simultaneously, but independently) that the complete network of FTAs, which corresponds to *efficient* global free trade, is pairwise stable; at the same time, they have also shown that some other *inefficient* configurations of FTA networks, in which one country is isolated and all the other form as many bilateral FTAs as possible, can be pairwise stable.

In their coalition/network formation approach, we find two common features embedded in their models. One is concerned with the *myopia* of the countries. In the coalition formation models, countries are considered to play the Nash equilibrium strategies in one-shot games. Although the concepts of SPNE and CPNE take account of the *farsightedness* on the side of the countries to some extent, they are myopic concepts and fail to capture the farsightedness of the countries satisfactorily. As for the network formation models, the pairwise stability requires for a particular network to be pairwise stable such that each country has no incentive to abandon an existing FTA and any pair of countries with no FTA between them has no incentive to form a new FTA. In the definition of pairwise stability, countries are supposed to only consider the immediate consequence of their actions of forming an FTA or abandoning an existing FTA, but ignore possible subsequent reactions by other countries. The notion of pairwise stability fails to deal with the farsightedness of the countries, too.

The other common feature is concerned with the ability of the models to predict the realization of the equilibrium outcomes. The solution concepts adopted in the coalition/network formation approach only tell us that *once an equilibrium FTA structure has been reached*, then no country wants to change it. These solution concepts together with the construction of the models, however, are totally silent about *whether and how an equilibrium FTA structure can be reached* from non-equilibrium FTA structures. In particular, these models, at the very outset, ignore the possibility that some interim FTAs might be formed en route to the final outcome.

The models based on the "dynamic" game approach have taken explicit account of the possibility of forming interim FTAs. Aghion, Antràs, and Helpman (2007) have considered a transferable utility, extensive form, bargaining game among three countries. In their model, one country is assumed to be the agenda-setter and others are assumed to be followers. The agenda-setter decides whether to consider sequential bilateral bargaining with other countries or to consider simultaneous multilateral bargaining with all countries at once, and also decides endogenously how much income transfers should be made. They have shown that if global free trade is efficient—if the *grand-coalition superadditivity* in the terminology of Aghion et al. (2007) is satisfied—and if the formation of bilateral FTAs imposes *negative externalities* on the outsider countries, then the agenda-setter prefers sequential bargaining and the grand coalition forms; in other words, some interim FTA between the agenda-setter and a follower country is formed on the SPNE path, and eventually, global free trade is reached.

Macho-Stadler and Xue (2007) have considered a transferable utility, partition function form, infinite horizon game among three countries. In their model, unlike in Aghion et al. (2007), countries are assumed to be *ex ante* symmetric and they move successively according to a predetermined order. Further, they assume that the surplus accruing from the formation of a trading bloc (CU) is divided through income transfers among the concerned countries according to a fixed sharing rule. Similar to Aghion et al. (2007), their model exhibits the grand-coalition superadditivity and the negative externality on the outsider countries. Then, they have shown that in the Markov Perfect Equilibrium (MPE) of the game, first, some trading bloc is formed, and eventually, a worldwide trading bloc is formed. Seidmann (2009) has, in a sense, extended the model of Macho-Stadler and Xue (2007) by allowing countries to form not only CUs but also FTAs. He has shed new light on a motive for forming a trading bloc, which he called the "strategic positioning": countries form a trading bloc in order to shift the status quo in a direction that is more favorable for member countries than the initial position.

In the models of Aghion et al. (2007) and Macho-Stadler and Xue (2007), the grand-coalition superadditivity and the negative externality due to the formation of trading blocs imposed on the outsider countries play important roles in achieving global free trade. The intuition behind this is explained as follows. If two or more countries form a trading bloc, the outsider countries become worse-off by the negative externalities; this gives the outsider countries higher incentives to form new trading blocs or to join the existing ones. Furthermore, the grand-coalition superadditivity guarantees (or, at least, makes it possible) that every country will be made better-off after global free trade is achieved. This explanation, however, relies on the fact that the existing trading blocs will never be dissolved as assumed (implicitly or explicitly) in Aghion et al. (2007), Macho-Stadler and Xue (2007), and Seidmann (2009). If countries can dissolve some of the existing trading blocs, even the "positive" externality can serve as a device that enhances the incentives of the countries to move toward global free trade. The point is not whether the externality is positive or negative, but that countries can change their positions strategically, and thereby, alter the incentive structure of the countries involved as pointed out by Seidmann (2009).

In this paper, as in Furusawa and Konishi (2005, 2007) and Goyal and Joshi (2006), we address the "dynamic time-path" question within the framework of the network formation games. Our approach, however, is different from theirs in that we take full account of the *farsightedness* on the side of the countries (players) and that we adopt the *von Neumann-Morgenstern (vNM) stable set*—the set of outcomes that satisfies both internal and external stabilities—as the solution concept.[5] With the notion of the vNM stable set, we can incorporate the farsightedness of the countries appropriately into the model. Further, the vNM stability (in particular, external stability) consistently explains whether and how some stable networks can be realized through the behavior of the farsighted players from other unstable networks; in other words, it takes account of the possibility of forming interim trading blocs as in Aghion et al. (2007), Macho-Stadler and Xue (2007), and Seidmann (2009). To highlight the roles of farsightedness and the notion of vNM stability, we also consider cases where countries are myopic and compare the vNM stable set with another well-known solution concept, that is, the core.

To make the model tractable, we adopt the same background trade model as Goyal and Joshi (2006) and Macho-Stadler and Xue (2007). We assume that there are *ex ante* symmetric countries; each country has one oligopolistic firm who sells a homogeneous good in both the domestic market and the foreign markets. Firms compete in the Cournot fashion. The markets are separated. Our model exhibits

[5]The notion of the vNM stable set was originally introduced by von Neumann and Morgenstern (1944) as a solution concept for games in characteristic function form. Greenberg (1990)'s theory of social situations has paved the way for its application to the games in other forms.

the grand-coalition superadditivity and the negative externality. Further, unlike Aghion et al. (2007), Macho-Stadler and Xue (2007), and Seidmann (2009), we assume away international income transfers (even within the member countries of a trading bloc).

Trading blocs in our model take the form of an FTA rather than that of a CU. That is, when a pair of countries form a bilateral FTA, they eliminate tariffs on mutual trade, but they do not coordinate the tariff rates on the imports from outsider countries. A pair of countries can form a new FTA without consent from the member countries of the existing FTAs; in addition, each single country can unilaterally annul the existing FTAs as many as it wants.

We assume that the pre-agreement tariff rates are given *exogenously*.[6] Then, we consider two extreme cases: one in which the pre-agreement tariffs are very (prohibitively) high and the other in which they are very low (almost zero). When the pre-agreement tariffs are very high, we show that (i) if the countries are myopic, there exists a unique nonempty myopic core consisting not only of the complete FTA network, which corresponds to global free trade, but also of FTA networks in each of which all countries except for one form a complete FTA network (i.e., a free-trade club) and the remaining one country is isolated—the myopic core in this case coincides with the set of all pairwise stable FTA networks; and that (ii) if the countries are farsighted, there exists a unique farsighted vNM stable set consisting only of the complete FTA network, which coincides with the farsighted core. On the other hand, when the pre-agreement tariffs are very low, we show that (iii) if the countries are myopic, there exists a unique nonempty myopic core consisting only of the complete FTA network; and that (iv) if the countries are farsighted, the farsighted core is empty, but (v) there exists a unique farsighted vNM stable set consisting of FTA networks in each of which all countries except for one form a free-trade club.

In the case of high pre-agreement tariffs, the farsighted vNM stable set, refining the myopic core and the set of pairwise stable networks, supports only global free trade. The external stability of the farsighted vNM stable set explains whether and how global free trade is achieved from other situations through the successive

[6]This assumption conforms Article XXIV of the GATT, which requires members of a preferential trade agreement (FTA in this case) to not raise tariffs on nonmembers.

formation (and/or dissolution) of bilateral FTAs. In this case, we can say that bilateral FTAs can be *building blocks* toward global free trade. On the other hand, in the case of low pre-agreement tariffs, the farsighted vNM stable set predicts that global free trade cannot be achieved; instead, it supports some other *inefficient* FTA networks. In this case, bilateral FTAs can be said to be *stumbling blocks* against global free trade.

Based on the same background trade model of Goyal and Joshi (2006) as the current paper, Xue and Zhang (2009) have examined an FTA network formation game incorporating the farsightedness of the countries. To capture the farsightedness, they have adopted the notion of *pairwise farsightedly stable set* developed by Herings, Mauleon, and Vannetelbosch (2009), which is closely related to (but, different from) our farsighted vNM stable set. The farsighted vNM stability requires both internal and external stabilities. On the other hand, the pairwise farsighted stability requires deterrence of external deviations, minimality, and external stability. Xue and Zhang (2009) have assumed, unlike our model, that the countries impose optimal tariffs on non-FTA countries. Their model admits a multiplicity of the pairwise farsightedly stable sets: one of them supports global free trade as a stable outcome, while the others do not. In contrast to this, the farsighted vNM stable set in our model is determined uniquely. The differences between their results and ours are attributable to the differences in the solution concepts.[7]

The rest of the paper is organized as follows. In Section 2, we introduce the background trade model and show how to represent FTA network configurations by using some graph-theoretical concepts. In Section 3, we explicitly formulate the FTA network formation game and introduce the solution concepts of the vNM stable set and the core. In Sections 4 and 5, we examine the high pre-agreement tariff case and the low pre-agreement tariff case, respectively, and show our main results. Section 6 contains some remarks.

To facilitate the discussion, all the proofs of the Lemmas and Theorems are relegated to the Appendices. Although we describe the model and state the lemmas and theorems based on the general n-country model as far as possible, some of the lemmas and theorems (in particular, Theorem 2 and Theorem 5) can be proved

[7]We will discuss this point later.

rigorously only in the benchmark model with four countries. If the restriction on the number of the countries is necessary, we state it explicitly.

2 Basic model

In each country, there is one oligopolistic firm, which can sell in the domestic market and the foreign markets. The markets are separated. If two countries have agreed on a bilateral FTA, then each country allows the other country's firm to enter its own market without imposing an import tariff. Otherwise, each country imposes a nonzero tariff on the imports from the other. For a firm of country k (i.e., firm k), the competitiveness and/or the profitability in country j's market depends not only on whether country k has established an FTA with country j but also on whether country j has established FTAs with other countries. This implies that the welfare of a country (which is the sum of the consumers' surplus, the profit of its firm, and the tariff revenue) depends on the whole structure of FTA configurations.

2.1 Demand, production, and welfare

Let $N = \{1, 2, \ldots, n\}$ be the set of symmetric countries. (As mentioned in the Introduction, we describe the model and state the lemmas and theorems in terms of the general n-country model as far as possible. When we intend to be more specific, we make use of the benchmark model with four countries.) In country j, there is a single firm (firm j) producing a homogeneous good with constant marginal cost technology; we assume that the marginal costs are zero. Let Q_j^k be the output of firm k in country j. Then, the total output supplied to country j's market is $Q_j \equiv \sum_{k\in N} Q_j^k$. Let p_j be the price of the good in country j. The inverse demand function of country j is given by $p_j = \alpha - Q_j$, where $\alpha > 0$.

The unit tariff rate faced by firm k in country j is denoted by t_k^j. The profit of firm k obtained from operating in country j is

$$\pi_j^k \equiv \left[\alpha - \sum_{i\in N} Q_j^i - t_k^j\right] Q_j^k. \tag{1}$$

Of course, the total profit of firm k is $\sum_{j\in N} \pi_j^k$. Firms compete in a Cournot fashion in each market. Because we assumed zero marginal costs, we can consider each country's market separately. The first-order condition for profit maximization of firm k in country j is

$$\alpha - \sum_{i \in N} Q_j^i - t_k^j - Q_j^k = 0. \tag{2}$$

By symmetry, we obtain the following result:

$$Q_j^k = \frac{\alpha - (n+1)t_k^j + \sum_{i \in N} t_i^j}{n+1}. \tag{3}$$

Then, the profit of firm k obtained from operating in country j can be expressed as follows:

$$\pi_j^k \equiv \left[\frac{\alpha - (n+1)t_k^j + \sum_{i \in N} t_i^j}{n+1}\right]^2. \tag{4}$$

On the other hand, the consumers' surplus CS^k of country k depends on the total supply in country k:

$$\mathrm{CS}^k = \frac{1}{2}\left[\frac{n\alpha - \sum_{i \in N} t_i^k}{n+1}\right]^2. \tag{5}$$

The tariff revenue R_j^k that country k can collect from firm j operating in country k's market is

$$R_j^k = t_j^k Q_k^j = \frac{\alpha t_j^k - (n+1)\left(t_j^k\right)^2 + \left(\sum_{i \in N} t_i^k\right) t_j^k}{n+1}. \tag{6}$$

Let N_k be the set of country k itself and the countries with which country k has FTAs; let n_k be the number of those countries, that is, $n_k \equiv |N_k|$. If country k and country j form an FTA, then they eliminate their mutual tariffs: $t_j^k = t_k^j = 0$. On the other hand, if country k has no FTA with country j, it sets its tariff rate at $t_j^k > 0$. Further, we assume that country k's tariff rates imposed on the imports from non-FTA countries are the same: $t_j^k = t^k > 0$ for all $j \in N \setminus N_k$. This reflects the Most-Favored Nations principle in the GATT/WTO.

The social welfare W^k of country k is measured by the sum of the consumers' surplus, the total tariff revenue, and the total profit of firm k:

$$\begin{aligned} W^k &= \mathrm{CS}^k + \sum_{j \in N} R_j^k + \sum_{j \in N} \pi_j^k \\ &= \frac{1}{2}\left[\frac{n\alpha - (n-n_k)t^k}{n+1}\right]^2 + \frac{t^k(n-n_k)\left\{\alpha - (1+n_k)t^k\right\}}{n+1} \\ &\quad + \sum_{j \in N_k}\left[\frac{\alpha + (n-n_j)t^j}{n+1}\right]^2 + \sum_{j \in N \setminus N_k}\left[\frac{\alpha - (1+n_j)t^j}{n+1}\right]^2. \end{aligned} \tag{7}$$

In general, the welfare of country k depends upon (i) the tariff rates of all countries (i.e., $t^1, t^2, \ldots, t^n$); (ii) the number of FTAs that country k forms (i.e., n_k); (iii) the number of FTAs that country k's partners have (i.e., n_j for $j \in N_k$); and (iv) the number of FTAs that non-partners have (i.e., n_j for $j \in N \setminus N_k$).

As is easily seen from Eq. (7), an increase in the tariff rate of a partner country (t^j for $j \in N_k$) raises the welfare of country k, while an increase in the tariff rate of a non-partner country (t^j for $j \in N \setminus N_k$) reduces the welfare of country k. Further, if a partner country or a non-partner country forms a new FTA, which increases the number of FTAs (n_j for $j \neq k$), then the welfare of country k deteriorates. In this sense, formation of an FTA in our model imposes "negative externalities" on the outsider countries.

2.2 Some graph-theoretical concepts

To describe FTA networks in a formal way, we need to introduce some graph-theoretical concepts.[8] Let V be a set of **vertices**. An unordered pair of distinct vertices is called an **edge**. Let E be a subset of the set of all edges. Then, the pair of sets (V, E) is called a **graph**.[9] For a graph G, we write $V(G)$ and $E(G)$ for its vertex set and edge set, respectively: that is, $G = (V(G), E(G))$. The **order** of a graph G is the number of vertices and the **size** of G is the number of edges. We denote the size of G as $e(G) \equiv |E(G)|$. An edge (v, w) is said to join the vertices v and w, and is simply denoted as vw. By definition, vw and wv represent the same edge. If $vw \in E$, then v and w are the **end-vertices** of vw and said to be **adjacent** to each other. If the context is clear, we simply write $v \in G$ and $vw \in G$ instead of writing $v \in V(G)$ and $vw \in E(G)$. The **degree** of a vertex v in G, denoted by $d_G(v)$, is the number of vertices adjacent to v in G. A vertex of degree zero is said to be **isolated**. An **empty graph** is a graph such that every vertex is isolated. A **complete graph** is a graph such that for every pair $\{v, w\}$ of distinct vertices, there

[8]Most of the definitions of the graph-theoretical concepts introduced here are borrowed from Bollobás (1979).

[9]The vertex set V is assumed to be nonempty, but the edge set E may be empty.

is an edge joining them.[10] The following fact is well-known and easy to prove:[11] for any graph G,

$$\sum_{v \in V(G)} d_G(v) = 2e(G). \tag{8}$$

A **walk** from v_0 to v_ℓ (i.e., a v_0-v_ℓ walk) in a graph G is an alternating sequence of vertices and edges $W = v_0, v_0v_1, v_1, v_1v_2, \ldots, v_{\ell-1}v_\ell, v_\ell$, where $v_{k-1}v_k \in E(G)$ for $k = 1, 2, \ldots, \ell$. The length of a v_0-v_ℓ walk is ℓ. A v_0-v_ℓ walk is called an **ℓ-cycle** if $\ell \geqq 3$, $v_0 = v_\ell$, and the vertices v_k ($0 \leqq k < \ell$) are all distinct. A **path** is a walk with distinct vertices. A graph is **connected** if for every pair $\{v, w\}$ of distinct vertices, there is a v-w path.

A graph (V_1, E_1) is a **subgraph** of another graph (V_2, E_2) if $V_1 \subset V_2$ and $E_1 \subset E_2$. A maximal connected subgraph of a graph G is a **component** of G.[12] A graph G can be divided into some components. Two graphs (V_1, E_1) and (V_2, E_2) are **isomorphic** if there exists a bijection $\varphi : V_1 \to V_2$ such that $vw \in E_1$ if and only if $\varphi(v)\varphi(w) \in E_1$. Intuitively speaking, two isomorphic graphs are of *the same shape*.

Given a graph $G = (V(G), E(G))$, we can obtain new graphs from G by deleting or adding some edges. If $(v, w) \in E(G)$, then $G - (v, w)$, or more simply, $G - vw$ denotes the graph obtained by deleting the edge (v, w) from G, that is, $G - vw = (V(G), E(G) \setminus \{(v, w)\})$. Similarly, if $(v, w) \notin E(G)$, then $G + (v, w)$ or $G + vw$ denotes the graph obtained by adding the edge (v, w) to G, that is, $G + vw = (V(G), E(G) \cup \{(v, w)\})$.

2.3 Representation of FTA networks

An FTA network is a description of how the countries form bilateral FTAs with other countries. An FTA network is represented by a graph G such that $V(G) = N$ and $(i, j) \in E(G)$ if and only if country i and country j have a bilateral FTA between them. Let Γ be the set of all graphs with their vertex sets being identified

[10] Note that some authors adopt a slightly different definition of a graph, in which two or more distinct edges that join v and w (i.e., parallels) as well as an edge that joins a vertex to itself (i.e., a loop) are allowed. In our definition, no parallels and loops are allowed.

[11] Each edge has exactly two end-vertices. Then, if we add the degree over all vertices, a particular edge is counted exactly twice. This proves Eq. (8).

[12] Here, "maximal" is taken with respect to the set inclusion.

with the set N of all countries. Then, Γ represents all possible FTA networks. The empty graph, denoted by $G^{\emptyset}$, corresponds to the situation where there is no FTA in the world. On the other hand, the complete graph, denoted by G^{*}, corresponds to the situation where every pair of countries forms a bilateral FTA—global free trade.

Let $N_k(G) \equiv \{k\} \cup \{j \in N | \ (k, j) \in E(G)\}$ be the set of country k itself and all countries with whom country k has FTAs in the FTA network G. Further, let $n_k(G) \equiv |N_k(G)|$ be the cardinality of $N_k(G)$. By definition, we have $n_k(G) = d_G(k) + 1 \geqq 1$ for all $k \in N$ and for all $G \in \Gamma$. By substituting $n_k(G)$ and $n_j(G)$ into Eq. (7), we can write the welfare of country k as a function of G, that is, $W^k(G)$ for all $G \in \Gamma$.

Examples of FTA networks in the benchmark model with four countries are illustrated in Figure 1. Small circles (i.e., a, b, c, and d) represent possible "addresses" at which countries are located (one address for one country) in a graph. The numbers in small circles represent the countries located at the corresponding addresses. A bold solid line between a pair of addresses means that a bilateral FTA is formed by the countries located at these addresses. On the other hand, a thin broken line between a pair of addresses means that there is no FTA between the countries located there.

Let G and H be the graphs that correspond to panel (i) and panel (ii) of Figure 1, respectively. Both G and H represent a situation where countries 1, 2, and 3 form a "hub-and-spoke" system of FTAs and the remaining country 4 is excluded from the hub-and-spoke system. Although the same three countries participate in the hub-and-spoke systems in both G and H, the welfare of each country obtained in G can be different from that in H. Take country 1 for example. Country 1 is a spoke in G, but it is the hub in H. We have $n_1(G) = 2$, $n_2(G) = 3$, $n_3(G) = 2$, and $n_4(G) = 1$ for G, while we have $n_1(H) = 3$, $n_2(H) = 2$, $n_3(H) = 3$, and $n_4(H) = 1$ for H. By substituting these results into Eq. (7), we can verify that $W^1(G) < W^1(H)$.[13] In general, the welfare of a country depends upon the current graph; more specifically, it depends both on the *shape* of the current graph (i.e., the isomorphic class to which the graph belongs) and on the *address* at which the

[13] See, also, Table 1.

country is located in the graph.

In the benchmark model with four countries, we have eleven different shapes of graphs. In other words, the set Γ of all possible graphs is partitioned into eleven isomorphic classes: $\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \Gamma_5, \Gamma_6, \Gamma_7, \Gamma_8, \Gamma_9, \Gamma_{10}$, and Γ_{11} such that $\Gamma = \bigcup_{r=1}^{11} \Gamma_r$ and $\Gamma_r \cap \Gamma_s = \emptyset$ if $r \neq s$. The shape of a (representative) graph in each isomorphic class Γ_r is illustrated in panel (r) of Figure 2 ($r = 1, 2, \ldots, 11$).

Let us examine some of the isomorphic classes. The isomorphic class Γ_1 contains only one graph, that is, the empty graph $G^{\emptyset}$. The empty graph $G^{\emptyset}$ means that there is no FTA in the world. The isomorphic class Γ_3, which contains six graphs, corresponds to a situation where the world is divided into two small trading blocs of equal size. For example, let us consider a graph $G \in \Gamma_3$ such that $E(G) = \{(1,2),(3,4)\}$ (i.e., $a = 1, b = 2, c = 3, d = 4$). This describes the situation where countries 1 and 2 form a bilateral FTA and countries 3 and 4 form another bilateral FTA and there are no other bilateral FTAs. The isomorphic class Γ_4 corresponds to a three-country hub-and-spoke system (exclusive of one isolated country). The isomorphic class Γ_5 contains four graphs. Let us consider a graph $G' \in \Gamma_5$ such that $E(G') = \{(1,2),(2,3),(3,1)\}$. This describes the situation where three countries 1, 2, and 3 form a free-trade club among them, but they exclude country 4 from the club. FTAs among countries 1, 2, and 3 constitute a complete component of G' (a 3-cycle graph). The isomorphic class Γ_7, which contains four graphs, corresponds to a worldwide hub-and-spoke system with one country being the hub and the other countries being spokes. Last, let us consider the isomorphic class Γ_{11}. This is a singleton consisting only of the complete graph G^*, which corresponds to global free trade.

3 Formation and stability of FTA networks

In the last section, we have considered how to describe the structure of FTA networks and their welfare implications. Here, we consider how an FTA network is formed and/or dissolved.

3.1 Inducement correspondence

Let G be a current FTA network. First, let us consider what a single country can do. A single country, of course, cannot form an FTA by itself, but it can annul the existing FTAs unilaterally. Suppose that country k has an FTA with country j ($j \in N_k(G) \setminus \{k\}$) in the current FTA network G. If country k annuls the existing FTA with country j, the current FTA network G changes to $G - kj$. Further, country k can choose some (possibly all) partner countries in G and cancel the FTAs with those countries simultaneously. Technically, from the current FTA network G, country k can *induce* another FTA network $H = G - \sum_{j \in T} kj$ for some $T \subset N_k(G) \setminus \{k\}$.[14] When a single country k can induce H from G in this way, we write $G \xrightarrow{\{k\}} H$ or $G \xrightarrow{k} H$.

Next, let $S = \{k, j\}$ be a coalition of two countries k and j and let us consider what S can do. If countries k and j have no FTA between them in G (i.e., $kj \notin G$) and if they negotiate and agree on a bilateral FTA, then a new FTA kj is added to the current network G, and accordingly, a new FTA network $G + kj$ is established. More generally, when countries k and j form a new FTA, country k, at the same time, can annul some of its FTAs with other countries unilaterally; the same is true for country j. In this case, from the current G, a two-country coalition S can induce $H = G + kj - \sum_{i \in T_k} ki - \sum_{i \in T_j} ji$ for some $T_k \subset N_k(G) \setminus \{k\}$ and $T_j \subset N_j(G) \setminus \{j\}$. If countries k and j have a bilateral FTA in G (i.e., $kj \in G$), countries k and j can annul some of the existing FTAs (possibly, including the FTA kj itself) at the same time. Therefore, in this case, S can induce $H = G - \sum_{i \in T_k} ki - \sum_{i \in T_j} ji$ for some $T_k \subset N_k(G) \setminus \{k\}$ and $T_j \subset N_j(G) \setminus \{j\}$. In this way, any coalition S of two countries can *induce* a new FTA network H from a current FTA network G. When a coalition $S = \{k, j\}$ of two countries k and j can induce H from G, we write $G \xrightarrow{S} H$ or $G \xrightarrow{\{k,j\}} H$.[15]

When there is a single country or a pair of countries that can induce H from G, we simply write $G \rightarrow H$. Following the terminology of the theory of social situations (Greenberg, 1990), we call the binary relation $\{\rightarrow\}$ defined on the set Γ of all graphs as the **inducement correspondence**. The inducement correspondence

[14] When $T = \{j_1, j_2, \ldots, j_R\}$, the expression $G - \sum_{j \in T} kj$ means $G - kj_1 - kj_2 - \cdots - kj_R$.

[15] A similar argument can also be applied to any nonempty coalition with an arbitrary number of countries.

only describes how countries *can* change the current FTA network to another network. It should be noted that neither $G \xrightarrow{\{k,j\}} H$ nor $G \xrightarrow{\{k\}} H$ implies that H is better than G for the countries concerned.

3.2 Domination relations

Let G be a current FTA network and consider a coalition $S = \{k, j\}$ of two countries. Suppose that $G \xrightarrow{S} H$. That is, the coalition S has some power to change the current FTA network G to another network H. However, whether the coalition S actually exercises this power depends on the welfare consequence and also on the *perspectives* of the countries in S as to how the other countries react to their initial actions. If the countries anticipate that exercising the power to induce H from G eventually leads to a situation in which they are made worse-off than the current situation, then they will not do so. On the other hand, if the countries anticipate that exercising the power eventually leads to a situation in which they are made better-off, then they do exercise their power.[16]

Taking different levels of the countries' perspectives into account, we define two domination relations on the set Γ of all possible FTA networks. The first one reflects the *myopia* of the countries.

Definition 1 (Direct domination relation). *For two FTA networks G and H, if there exists a nonempty coalition S of countries with $1 \leqq |S| \leqq 2$ such that $G \xrightarrow{S} H$ and $W^i(G) < W^i(H)$ for all $i \in S$, then we say that "H directly dominates G through S" and we write $G \prec_S H$. Further, when $G \prec_S H$ for some nonempty $S \subset N$, we simply say that "H directly dominates G" and write $G \prec H$.*

Consider a coalition S that *can* induce H from G (i.e., $G \xrightarrow{S} H$). If all the countries in S believe that they will be made better-off in H than in G and also believe that H will remain in the status quo after being induced (i.e, S believes that the other countries do not react in H at all), the countries actually exercise their power to induce H from G. Thus, the direct domination relation $G \prec_S H$ is realized if the countries in S ignore the possibility that H will be replaced with another network H' through subsequent (re)actions by other countries. In this

[16]The same explanation also applies to a single country.

sense, the direct domination relation reflects the *myopia* of the countries in S. When $G \prec_S H$, we say that the countries in S have *myopic incentives* to move from G to H.

The indirect domination relation defined below reflects the *farsightedness* of the countries.

Definition 2 (Indirect domination relation). *For two FTA networks G and H, if there exist a sequence of FTA networks $\{G_r\}_{r=0}^R$ and a corresponding sequence of coalitions of countries $\{S_r\}_{r=1}^R$ that satisfy the following conditions:*

(i) *$G_0 = G$ and $G_R = H$,*

(ii) *$G_{r-1} \xrightarrow{S_r} G_r$ and $1 \leqq |S_r| \leqq 2$ for all $r = 1, 2, \ldots, R$, and*

(iii) *$W^i(G_{r-1}) < W^i(H)$ for all $i \in S_r$ and for all $r = 1, 2, \ldots, R$,*

then we say that "H indirectly dominates G" and we write $G \ll H$.

The second requirement that the number of deviating countries in each step of the indirect domination should not exceed two reflects the *bilateralism* in our model. If we relax this requirement and allow the simultaneous deviation by more than three (possibly, by all) countries, we can define another indirect domination relation that reflects the *multilateralism*. After showing our main results, we will discuss how the multilateralism works in our model.

Let us consider a coalition S_r that appears in the definition of the indirect domination relation. When S_r induces G_r from G_{r-1}, country k in S_r compares the current welfare level $W^k(G_{r-1})$ with the welfare level $W^k(H)$ that can be obtained in the "final" FTA network H, but not with the welfare level $W^k(G_r)$ that can be obtained in the "immediate" FTA network G_r after G_{r-1}. Therefore, it may be that S_r *does* exercise the power to change G_{r-1} to G_r even if some countries in S_r are made worse-off in G_r immediately after G_{r-1}, anticipating that subsequent inducements by other countries will eventually lead to H in which all the countries in S_r are made better-off than in G_{r-1}. In this sense, the indirect domination relation reflects the *farsightedness* of the countries.[17]

[17] Note that $G \prec H$ implies $G \ll H$.

The definition of indirect domination relation is akin to that of a *farsighted improving path* introduced by Herings et al. (2009) in order to define their solution concept of the pairwise farsightedly stable set. In each step of a farsighted improving path, only one link (edge) can be formed or severed. On the other hand, in each step of a sequence realizing indirect domination, only one link (edge) can be formed, but, at the same time, some of the existing links can be severed (if a country or a pair of countries concerned wants to do so). In a sense, the indirect domination relation allows the countries a higher degree of freedom to deviate from a current FTA network than the farsighted improving path. This fact plays an important role in establishing our main results.[18]

3.3 Solution concepts

In general, an **abstract system** is a pair $(X, \angle)$ of a nonempty set X and a domination relation $\angle$ defined on X: for $x, y \in X$, we say that y dominates x if $x \angle y$. Based on the domination relations defined in the previous subsection, we have two particular abstract systems $(\Gamma, \prec)$ and $(\Gamma, \ll)$, which we call the **myopic system** and the **farsighted system**, respectively. Let us define the solution concepts for $(X, \angle)$.

Definition 3 (Core). *The core of an abstract system $(X, \angle)$, denoted by $C(X, \angle)$, is a subset of X consisting of all outcomes that are not dominated. That is,*

$$C(X, \angle) \equiv \{x \in X| \text{ there is no } y \in X \text{ such that } x \angle y\}. \tag{9}$$

The core of $(X, \angle)$ always exists and is unique, but it can be an empty set. Once an outcome in the core has been reached, it will never be replaced with any other outcome according to the domination relation $\angle$. Therefore, we can say that in a sense, outcomes in the core are "stable." The concept of the core, however, fails to explain whether and how outcomes in it can be reached from (unstable) outcomes outside of the core.

Definition 4 (von Neumann-Morgenstern stable set). *The vNM stable set is a subset K of X that satisfies the following conditions:*

[18]The proof of Theorem 5, for example, heavily relies upon the fact that a pair of two countries can form an FTA between them and, at the same time, abondon some of the existing FTAs. See the discussion after Theorem 5.

(i) *for all* $x \in K$*, there does not exist* $y \in K$ *such that* $x \angle y$ (**internal stability**)*;*

(ii) *for all* $x \in X \setminus K$*, there exists* $y \in K$ *such that* $x \angle y$ (**external stability**).

The set of all vNM stable sets for $(X, \angle)$ *is denoted by* $\mathcal{K}(X, \angle)$.

Let K be a vNM stable set for $(X, \angle)$. Internal stability means that any stable outcome, in the sense that it is included in K, will not be replaced with any other stable outcome according to $\angle$. On the other hand, external stability means that any unstable outcome, in the sense that it is excluded from K, will be replaced with some stable outcome according to $\angle$. In this manner, the concept of the vNM stable set (in particular, its external stability) consistently explains what the concept of the core fails to do.

We call the core and a vNM stable set for the myopic system $(\Gamma, \prec)$ as the **myopic core** and a **myopic vNM stable set**, respectively; similarly, we call those for the farsighted system $(\Gamma, \ll)$ as the **farsighted core** and a **farsighted vNM stable set**, respectively. Although the definitions are slightly different, the myopic core in our model coincides with the set of all *pairwise stable* networks as defined by Jackson and Wolinsky (1996).

Based on the notion of farsighted improving path, Herings et al. (2009) have defined the *von Neumann-Morgenstern (vNM) pairwise farsightedly stable set* and examined its relation to the pairwise farsightedly stable set. They have shown, in particular, that a vNM pairwise farsightedly stable set is a pairwise farsightedly stable set. The difference in the definitions of the farsighed improving path and the indirect domination relation renders our farsighted vNM stable set different from the vNM pairwise farsightedly stable set. The farsighted vNM stable set is not necessarily a pairwise farsightedly stable set; therefore, the predictions by the farsighted vNM stable set and by the vNM pairwise farsightedly stable set can be different.

4 High pre-agreement tariff case

We first consider the case in which the pre-agreement tariff rates are very high. More specifically, we assume, as in Goyal and Joshi (2006), that each country levies prohibitive tariffs on the imports from countries with whom the country does not have bilateral FTAs. In this case, international trade occurs only between those countries who have bilateral FTAs; isolated countries (in a graph-theoretical sense) actually adopt isolationist policies (i.e., they are in autarky). Therefore, no country earns tariff revenue. Then, the welfare of country k in an FTA network G can be expressed as the sum of the consumers' surplus and the profits obtained in the markets of country k's partners. Accordingly, Eq. (7) is reduced to

$$W^k(G) = \frac{1}{2}\left[\frac{\alpha n_k(G)}{n_k(G)+1}\right]^2 + \sum_{j \in N_k(G)} \left[\frac{\alpha}{n_j(G)+1}\right]^2. \tag{10}$$

Note that the negative externality on the welfare of country k in this case only comes from an increase in the number of FTAs that country k's partners have (i.e., the number of partners' partners). The reason is quite simple. If country k's partner (say, country j) forms a new FTA with a third country (say, country i), then the market in country j becomes more competitive due to the entry of firm i after the formation of the FTA (j, i). Because firm k has no access to country i's market, it only suffers from a decrease in its profit obtained in country j's market. This makes country k worse-off. On the other hand, even if countries separated from country k and its partners form some new FTAs, the markets in which firm k is operating are not affected by the formation of these new FTAs. Firm k experiences no loss (gain) in its profit, and therefore, the welfare of country k is kept unchanged.

Because Eq. (10) contains only one exogenous parameter (i.e., α), we can easily calculate the value of $W^k(G)$ for every $G \in \Gamma$. As mentioned before, country k's welfare depends both on the isomorphic class to which G belongs and on the address at which country k is located in G. Table 1 summarizes the values of $W^k(G)$ in our benchmark model with four countries. The table also shows world welfare, which is defined as the sum of the welfare of all countries. As can be seen from Table 1, a single country can attain the highest welfare when it becomes the hub of a worldwide hub-and-spoke system (i.e., when it is located at b in a graph $G \in \Gamma_7$), while world welfare is maximized when global free trade is achieved (i.e., $G^* \in \Gamma_{11}$). In our model, the complete network G^* is *efficient*, and therefore, satisfies the *grand-coalition superadditivity*.[19]

Concerning the incentives of a country to form FTAs, Goyal and Joshi (2006) have shown the following results.

Observation 1 (Goyal and Joshi, 2006).

(i) *If a country is involved in one or more FTAs, it has an incentive to form an additional FTA.*

(ii) *If $n \geqq 4$, in an FTA network in which one country is isolated and the other $n-1$ countries constitute a complete component FTA network, the isolated country has no incentive to form an additional FTA.*

Based on the above observations, Goyal and Joshi (2006) have shown that (i) the complete network (i.e., global free trade) is pairwise stable, (ii) if $n \geqq 4$, an FTA in which one country is isolated and the other $n-1$ countries constitute a complete component FTA network (i.e., a free-trade club with $n-1$ countries) is pairwise stable, and (iii) there is no other type of pairwise stable networks. In our benchmark model with four countries, we can rephrase their results as follows:

Theorem 1. *In the case of high pre-agreement tariffs with four countries, the myopic core $C(\Gamma, \prec)$ is nonempty and is characterized by*

$$C(\Gamma, \prec) = \Gamma_5 \cup \Gamma_{11}. \tag{11}$$

The formal proof is omitted; instead, we give an illustration. Bold black arrows in Figure 3 represent the direct domination relation when the pre-agreement tariffs are high.[20] For example, the arrow from panel (2) to panel (3) means that for any FTA network $G \in \Gamma_2$, there exists an FTA network $H \in \Gamma_3$ that directly dominates G. Other bold black arrows in the figure carry similar information about the direct domination relation. It should be noted that any FTA network in some isomorphic class cannot directly dominate other FTA networks in the same

[19] The fact that the complete network G^* is efficient in a general n-country model has been proved by Goyal and Joshi (2006).

[20] For the moment, ignore other symbols such as thin arrows with white heads, double circles, and asterisks.

isomorphic class. As illustrated in Figure 3, there is no bold black arrow that begins from panel (5) or from panel (11), meaning that any FTA network in Γ_5 or Γ_{11} cannot be directly dominated by other FTA networks. This proves Theorem 1.

When the countries are myopic and the pre-agreement tariffs are very high, global free trade is *not* the only stable outcome. Once a free-trade club with $n-1$ countries (exclusive of one country) has been formed, the world is trapped in an inefficient situation. Although each member of the free-trade club has an incentive to form a new FTA with the isolated country, the isolated country has no *myopic incentive* to do so.

Let us consider, in the benchmark model, an FTA network $G \in \Gamma_5$ where there exist a three-country free-trade club and an isolated country. Suppose that countries 1, 2, and 3 form a free-trade club and country 4 is isolated in G. Because country 4 is located at d in G, it receives $W^4(G) = 3\alpha^2/8$ in G. On the other hand, a member of the free-trade club (say, country 3) receives $W^3(G) = 15\alpha^2/32$ in G. If global free trade G^* is achieved, both countries 3 and 4 will receive $W^3(G^*) = W^4(G^*) = 12\alpha^2/25$ (no matter where they are located in G^*), which is higher than $W^3(G)$ and $W^4(G)$. Therefore, if all countries are *farsighted* enough to understand the consequences of not only the immediate outcome of their own action but the final outcome that will be realized through the chain reactions of other countries, then, anticipating that global free trade will be realized eventually, countries 3 and 4 may form a new FTA between them (even though country 4 becomes worse-off in the immediate FTA network). In the following, we show that this is indeed the case.

Lemma 1. *In the case of high pre-agreement tariffs with four countries,*

(i) *no FTA network can indirectly dominate the complete network G^*;*

(ii) *the complete network G^* indirectly dominates any other FTA network.*

With the above results, we can establish the following theorem.

Theorem 2. *In the case of high pre-agreement tariffs with four countries, there exists a unique farsighted stable set K_H that is characterized by*

$$K_H = \Gamma_{11} \equiv \{G^*\}. \tag{12}$$

Further, K_H coincides with the farsighted core $C(\Gamma, \ll)$ of the farsighted system.[21]

Comparing Theorem 1 and Theorem 2, we find that FTA networks in the isomorphic class Γ_5, which are included in the myopic core, fail to be stable when the countries are farsighted. To see how FTA networks in Γ_5 are "destabilized" through the behavior of farsighted countries, let us consider the following example. Take an FTA network $G \in \Gamma_5$ such that $E(G) = \{(1,2),(2,3),(3,1)\}$ and country 4 is isolated. Consider the coalitions $\{1,4\},\{2,4\},\{3,4\} \subset N$; correspondingly, let us define the networks such that $G_1 = G + (1,4)$, $G_2 = G_1 + (2,4)$, and $G_3 = G_2 + (3,4)$. By definition and from Table 1, we obtain the following results:

$$\begin{array}{llll} G \xrightarrow{\{1,4\}} G_1 \in \Gamma_8 & \text{and} & W^i(G) < W^i(G^*) & \text{for} \quad i = 1,4, \\ G_1 \xrightarrow{\{2,4\}} G_2 \in \Gamma_{10} & \text{and} & W^i(G_1) < W^i(G^*) & \text{for} \quad i = 2,4, \\ G_2 \xrightarrow{\{3,4\}} G_3 = G^* \in \Gamma_{11} & \text{and} & W^i(G_2) < W^i(G^*) & \text{for} \quad i = 3,4. \end{array}$$

Therefore, the sequence $G \xrightarrow{\{1,4\}} G_1 \xrightarrow{\{2,4\}} G_2 \xrightarrow{\{3,4\}} G_3 = G^*$ realizes $G \ll G^*$. The key is the behavior of country 4 in the first step in this sequence: country 4 together with country 1 induces G_1 from G. In the first step, country 4 itself becomes worse-off. At the same time, due to the negative externality, countries 2 and 3 become worse-off; therefore, they have higher incentives to move from G_1 toward global free trade G^*. This makes it possible to realize global free trade.

Theorem 2 implies that if the pre-agreement tariffs are very high, global free trade is the only final outcome that emerges from the chains of bilateral FTA negotiations among the countries. Even under the prevalence of bilateralism (as embedded in the definition of our inducement correspondence), global free trade can be achieved through the formation of bilateral FTAs by farsighted countries. In this case, we can say that bilateral FTAs are *building blocks* toward global free trade.

[21] Applying the notion of the (myopic) vNM stable set to Krugman (1993)'s monopolistically competitive FTA formation model, Oladi and Beladi (2008) have shown that there exist a unique myopic vNM stable set supporting global free trade. Although it can be easily shown that there exists a unique myopic vNM stable set for our model, it contains not only global free trade but also some other inefficient FTA networks. We omit the proof of existence and the characterization of the myopic vNM stable set for our model, because it is of less analytical interest.

5 Low pre-agreement tariff case

In this section, we consider the case where the pre-agreement tariff rates are very low. In this case, even an isolated country (in a graph-theoretical sense) engages in international trade with all other countries. In other words, even if country k does not have FTAs with some (possibly, all) countries in the current situation, firm k is operating in the markets of those countries. The welfare of country k is represented by Eq. (7), which contains many exogenous parameters such as t^k, t^j, and α. To compare the welfare levels at different addresses in different graphs effectively, we assume that $t^k = t^j = t$ for all $k, j \in N$ in Eq. (7) and that t is positive but sufficiently close to zero.

The following lemma characterizes the differences in the welfare of a single country and in world welfare across different FTA networks.

Lemma 2. *In the case of low pre-agreement tariffs, we have for all $k \in N$ and for all $H, G \in \Gamma$,*

(i) $W^k(H) > W^k(G)$ *if and only if*

$$\Delta \equiv n_k(H) - n_k(G) - \frac{4}{2n+1}\{e(H) - e(G)\} > 0 \quad \textit{and} \tag{13}$$

(ii) $\sum_{k \in N} W^k(H) > \sum_{k \in N} W^k(G)$ *if and only if*

$$e(H) - e(G) > 0. \tag{14}$$

The difference in the welfare of a single country (say, country k) in different FTA networks depends both on the number of FTAs that country k has and on the number of existing FTAs in the world. On the other hand, the difference in world welfare in different FTA networks depends only on the number of existing FTAs in the world. As the number of FTAs increases, world welfare increases. The following results are immediate from Lemma 2.

Observation 2.

(i) *If FTAs not involving country k are formed, the welfare of country k decreases.*

(ii) *If FTAs involving country k and no other FTAs are formed, the welfare of country k increases.*

(iii) *Global free trade is efficient.*

Observation 2-(i) implies the *negative externality* due to the formation of FTAs. In the high pre-agreement tariff case, the negative externality on country k only comes from the formation of FTAs by the countries with whom country k already has formed FTAs. In the low pre-agreement tariff case on the other hand, even the formation of FTAs by the countries that are separated from country k in the current situation can negatively affect the welfare of country k. In this sense, we can say that the negative externality is stronger in the low pre-agreement tariff case than in the high pre-agreement tariff case.

Observation 2-(ii) contrasts with Observation 1-(ii). In the low pre-agreement tariff case, every country always has a *myopic incentive* to form as many new FTAs as possible. This observation leads us to the following result.

Theorem 3. *In the case of low pre-agreement tariffs, the myopic core $C(\Gamma, \prec)$ is nonempty and it is characterized by*

$$C(\Gamma, \prec) = \{G^*\}. \tag{15}$$

Note that the complete graph constitutes the myopic core in a general n-country model and that the complete graph is the only pairwise stable FTA network in this case. In Figure 3, the direct domination relation in our benchmark model is illustrated by thin arrows with white heads. For example, the arrow pointing from Γ_5 to Γ_8 implies that for any $G \in \Gamma_5$, there exists an FTA network in Γ_8 that directly dominates G. As illustrated in the figure, Γ_{11} is the only isomorphic class that has no originating arrow.[22]

If the countries are myopic and the pre-agreement tariffs are sufficiently low, global free trade G^* is the only stable outcome in the sense that it is included in the myopic core. From a single country's viewpoint, because global free trade contains as many FTAs as possible, the negative externality is maximized in G^*.

[22] Differences between the direct domination relation in the low pre-agreement tariff case and that in the high pre-agreement tariff case only appear in the relations between Γ_4 and Γ_7 and between Γ_5 and Γ_7.

Therefore, if the countries are farsighted, some of them may want to abandon some existing FTAs, anticipating the eventual realization of an FTA network in which they can be made better-off. The following lemma clearly shows this point.

Lemma 3. *In the case of low pre-agreement tariffs with four countries, global free trade G^* is indirectly dominated by an FTA network in* Γ_5.

The proof is omitted; see the discussion after Theorem 5. This lemma together with Observation 2-(ii) implies the following results.

Theorem 4. *In the case of low pre-agreement tariffs with four countries, the farsighted core $C(\Gamma, \ll)$ is empty.*

The proof is omitted; this follows directly from Theorem 3 and Lemma 3. When the countries are farsighted and the pre-agreement tariffs are sufficiently low, the notion of the core predicts nothing.

Myopic countries, only taking account of the immediate effects of their own FTA formation, do not want to leave global free trade. On the other hand, in order to avoid strong negative externalities, farsighted countries have incentives to leave global free trade and to induce other FTA configurations with less FTAs in the world. The tension between the incentives of a country to enlarge FTAs (Observation 2-(ii)) and to avoid the negative externality due to the formation of FTAs by other countries determines the stability of FTA networks.

Theorem 5. *In the case of low pre-agreement tariffs with four countries, there exists a unique farsighted stable set K_L that is characterized by*

$$K_L = \Gamma_5. \tag{16}$$

To show how global free trade G^* is "destabilized" and how an FTA network H in K_L is realized, we give an example of a sequence that realizes the indirect domination $G^* \ll H$. Let us consider the following set of FTA networks:

$$\begin{aligned}
&G_1 \in \Gamma_{10}: E(G_1) = \{(1,2),(1,3),(2,3),(2,4),(3,4)\}, && e(G_1) = 5,\\
&G_2 \in \Gamma_8: \ E(G_2) = \{(1,2),(2,3),(2,4),(3,4)\}, && e(G_2) = 4,\\
&G_3 \in \Gamma_6: \ E(G_3) = \{(1,2),(2,4),(3,4)\}, && e(G_3) = 3,\\
&G_4 \in \Gamma_9: \ E(G_4) = \{(1,2),(1,3),(2,4),(3,4)\}, && e(G_4) = 4,\\
&H \in \Gamma_5: \ E(H) = \{(1,2),(1,3),(2,3)\}, && e(H) = 3.
\end{aligned}$$

In H, countries 1, 2, and 3 form a free-trade club, but country 4 is excluded. As compared to global free trade G^*, each member of the free-trade club loses only one FTA in H, while the number of FTAs in the world is reduced by three. Each member experiences less negative externalities in H than in G^*. Country 1 can induce G_1 from G^* unilaterally by abandoning the FTA with country 4. Further, because $n_1(H) - n_1(G^*) = -1$ and $e(H) - e(G^*) = -3$, we have $W^1(H) > W^1(G^*)$ by Lemma 2. That is,

$$G^* \xrightarrow{\{1\}} G_1 \quad \text{and} \quad W^1(G^*) < W^1(H).$$

Similarly, we can show that

$$\begin{aligned}
&G_1 \xrightarrow{\{1\}} G_2 && \text{and} && W^1(G_1) < W^1(H),\\
&G_2 \xrightarrow{\{3\}} G_3 && \text{and} && W^3(G_2) < W^3(H),\\
&G_3 \xrightarrow{\{1,3\}} G_4 && \text{and} && W^i(G_3) < W^i(H) \quad \text{for} \quad i = 1,2,\\
&G_4 \xrightarrow{\{2,3\}} H && \text{and} && W^i(G_4) < W^i(H) \quad \text{for} \quad i = 2,3.
\end{aligned}$$

In the second step, country 1 cancels the FTA with country 3 unilaterally. (We can compress the first and second steps by assuming that country 1 annuls the FTAs with countries 3 and 4 simultaneously.) Then, country 2 comes to occupy a hub-like position in G_2. In the third step, country 3 cancels the FTA with country 2. Then, countries 1 and 3 form an FTA in the fourth step. In the last step, countries 2 and 3 form an FTA between them, and at the same time, abandon the FTAs with country 4 simultaneously. We obtain $G^* \ll H$.[23]

[23] This proves Lemma 3.

It is worth noting that in the above sequence realizing $G^* \ll H$, the FTA between countries 1 and 3 is abandoned once (by country 1 unilaterally) and is reorganized by them bilaterally. Similar argument also applies to the FTA between countries 2 and 3. In this way, by forming and dissolving FTAs, countries can change their relative addresses in FTA networks strategically. This can be seen as a reflection of the *strategic positioning* as indicated by Seidmann (2009).

Our Theorem 5 is in sharp contrast to the results obtained by Aghion et al. (2007) and Macho-Stadler and Xue (2007). Although their models and our model share important properties such as the grand-coalition superadditivity, negative externality, and farsightedness of the countries, their models predict the realization of global free trade, while our model predicts the realization of other inefficient FTA structures. The possibility of dissolving the existing FTAs and the strength of the negative externality are the differences between their models and ours. In their models, it is assumed that if FTAs have been formed, they will never be dissolved. On the other hand, in our model, it is assumed that countries can form and/or dissolve FTAs; as such, countries can make the most of their *strategic positioning* to avoid the strong negative externality accruing from the formation of FTAs by other countries. If we assume away the possibility of dissolving the existing FTAs from our model, we can realize global free trade, but we cannot exclude the inefficient FTA networks from the farsighted vNM stable set.[24]

Thus far, we have assumed that only a single country or a pair of two countries can induce one FTA network from another FTA network; in other words, we have only considered the role of *bilateralism*. Here, let us briefly discuss the role of *multilateralism*. By dropping the requirement of $|S_r| \leqq 2$ in the definition of the indirect domination relation $\ll$, we can define a new indirect domination relation under multilateralism. Is global free trade secured by the farsighted vNM stable set under the new indirect domination relation? Unfortunately, the answer is no. Even though the grand-coalition superadditivity is satisfied in our benchmark model, a coalition with three countries has a strong incentive to deviate from

[24]The inducement correspondence can be modified easily to not allow the possibility of dissolving the existing FTAs. Further, it is easy to show the existence of the farsighted vNM stable set, which includes global free trade and other inefficient FTA networks, under the modified inducement correspondence.

global free trade G^* to an FTA network in the isomorphic class Γ_5, and actually, it can do so in only one step. Global free trade is "destabilized" much easily under multilateralism than under bilateralism when the pre-agreement tariffs are sufficiently close to zero.[25]

6 Remarks

We have shown that when the pre-agreement tariffs are very high, the farsighted vNM stable set only supports global free trade, and that when the pre-agreement tariffs are sufficiently low, the farsighted vNM stable set does not support global free trade and instead supports some inefficient FTA networks. In the former case, we can say that bilateral FTAs are the *building blocks* for achieving global free trade, and in the latter case, to the contrary, they are the *stumbling blocks* against achieving global free trade. These results make a somewhat ironic impression: the closer the world economy is to global free trade (in the sense that the pre-agreement tariffs are very low), the harder it is to realizing the same.

Of course, our results depend upon the details and strong assumptions of our model. In particular, we have assumed that the tariff rates on non-FTA countries are exogenously determined. As Yi (2000) and Bond et al. (2004) have reported, if the tariff rates on non-FTA countries are determined *endogenously* to maximize the social welfare of each country, the tariff rates after the formation of an FTA decrease.[26] This *tariff-complementarity* effect makes non-FTA countries better-off, and thereby, generates *positive externalities* on them.[27] Then, the mechanisms supporting our results and the roles of bilateralism and multilateralism can be reversed. Hence, it may be the case that global free trade is supported by the farsighted vNM stable set. To investigate this possibility is a subject of future research.

[25]This conclusion depends upon the non-availability of international transfers in our model. If we allow international transfers *and* coalitional moves by more than three countries, the isolated country in a stable FTA network in Γ_5 becomes able to bribe all the other countries to move toward global free trade G^*. In this case, it may be possible (though not proved) that Γ_5 is destabilized and global free trade is established as the stable outcome through the behavior of farsighted countries.

[26]The fact that the optimal tariff rate after the formation of an FTA is lower than before has been reported by several other authors.

[27]The term "tariff-complementarity" is attributable to Bagwell and Staiger (1998).

A Appendix: Proofs

A.1 Proof of Lemma 1

Part (i)

Suppose, in negation, that there exists H that indirectly dominates G^* and that the following sequence of FTA networks $\{G_r\}_{r=1}^R$ and corresponding coalitions $\{S_r\}_{r=1}^R$ realizes $G^* \ll H$:

$$G^* \xrightarrow{S_1} G_1 \xrightarrow{S_2} G_2 \xrightarrow{S_3} \cdots \xrightarrow{S_R} G_R = H. \tag{17}$$

Note that because no pair of two countries can form a new FTA in G^*, the first coalition S_1 must dissolve one or some of the existing FTAs in G^*. Because the situation in G^* is symmetric for all countries, we can assume $S_1 = \{1\}$ or $S_1 = \{1, 2\}$. Further, without loss of generality, we can concentrate on the welfare of country 1. If G^* were to be indirectly dominated by H, we must have $W^1(G^*) < W^1(H)$. From Table 1, we can show that $W^1(G^*) < W^1(H)$ holds true only in the following five cases: (i) $H \in \Gamma_4$ and country 1 is located at b; (ii) $H \in \Gamma_6$ and country 1 is located at b or at c; (iii) $H \in \Gamma_7$ and country 1 is located at b; (iv) $H \in \Gamma_8$ and country 1 is located at b; and (v) $H \in \Gamma_{10}$ and country 1 is located at b or at d. Figure 3 illustrates the situation: vertices with asterisks ($*$) mean that countries located at these addresses receive higher welfare than when they are in G^*.

We prove only case (i), because essentially the same proof applies to the other cases as well. Because country 1 has three FTAs in G^*, it can annul one or two or three FTAs unilaterally. Therefore, we have three subcases: (i-1) $G_1 \in \Gamma_5$ (when country 1 annuls all three FTAs), (i-2) $G_1 \in \Gamma_8$ (when country 1 annuls two of the three FTAs), and (i-3) $G_1 \in \Gamma_{10}$ (when country 1 annuls one of its FTAs).

Let us consider subcase (i-1). In H, because country 1 is located at b, each of the other three countries must be located at a or c or d. Then, from Table 1, we have $W^i(H) = 19\alpha^2/48$ or $W^i(H) = 3\alpha^2/8$ for $i = 2, 3, 4$ in H. In G_1, on the other hand, country 1 is isolated and the other three countries form a 3-cycle network. From Table 1, we have $W^i(G_1) = 15\alpha^2/32$ for $i = 2, 3, 4$. Hence, we have $W^i(G_1) > W^i(H)$ for $i = 2, 3, 4$. This implies that countries 2, 3, and 4 will not be members of S_2 in sequence (17). Therefore, S_2 must be a singleton consisting only of country 1. Because country 1 is isolated in G_1, it cannot induce any other FTA network from G_1 unilaterally—a contradiction. Hence, subcase (i-1) is not possible.

Next, let us consider subcase (i-2). Country 1 is located at a and the other countries are located at b, c, or d in G_1. Similar to subcase (i-1), we have $W^i(G_1) > W^i(H)$ for $i = 2, 3, 4$. Then, countries 2, 3, and 4 will not be members of S_2. The only possibility is $S_2 = \{1\}$ and G_2 must be in Γ_5. Once an FTA network in Γ_5 has been reached, the same argument as subcase (i-1) applies. Hence, subcase (i-2) is not possible either.

Last, let us consider subcase (i-3). Country 1 is located at a or c in G_1. Similar to subcases (i-1) and (i-2), we must have $S_2 = \{1\}$ and $G_2 \in \Gamma_5 \cup \Gamma_8$. Once an FTA network in either Γ_5 or Γ_8 has been reached, the situation becomes parallel to subcases (i-1) or (i-2). Hence, subcase (i-3) is not possible either. This completes the proof of part (i).

Part (ii)

Take the empty graph $G^\emptyset \in \Gamma_1$. We now show that $G^\emptyset$ is indirectly dominated by the complete network G^*. By definition, we have $E(G^\emptyset) = \emptyset$ and $E(G^*) = \{(1,2), (2,3), (3,4), (4,1), (2,4), (1,3)\}$. To be specific, let us consider a sequence of FTA networks $\{G_r\}_{r=1}^5$ whose edge sets are given as follows:

$$
\begin{aligned}
E(G_1) &= \{(1,2)\},\\
E(G_2) &= \{(1,2), (3,4)\},\\
E(G_3) &= \{(1,2), (2,3), (3,4)\},\\
E(G_4) &= \{(1,2), (2,3), (3,4), (4,1)\},\\
E(G_5) &= \{(1,2), (2,3), (3,4), (4,1), (2,4)\}.
\end{aligned}
$$

By construction and from Table 1, we have

$$
\begin{array}{llll}
G^\emptyset \xrightarrow{\{1,2\}} G_1 \in \Gamma_2 & \text{and} & W^i(G^\emptyset) < W^i(G^*) & \text{for} \quad i = 1, 2,\\
G_1 \xrightarrow{\{3,4\}} G_2 \in \Gamma_3 & \text{and} & W^i(G_1) < W^i(G^*) & \text{for} \quad i = 3, 4,\\
G_2 \xrightarrow{\{2,3\}} G_3 \in \Gamma_6 & \text{and} & W^i(G_2) < W^i(G^*) & \text{for} \quad i = 2, 3,
\end{array}
$$

$$
\begin{aligned}
&G_3 \xrightarrow{\{4,1\}} G_4 \in \Gamma_9 \quad \text{and} \quad W^i(G_3) < W^i(G^*) \quad \text{for} \quad i = 1, 4, \\
&G_4 \xrightarrow{\{2,4\}} G_5 \in \Gamma_{10} \quad \text{and} \quad W^i(G_4) < W^i(G^*) \quad \text{for} \quad i = 2, 4, \\
&G_5 \xrightarrow{\{1,3\}} G^* \in \Gamma_{11} \quad \text{and} \quad W^i(G_5) < W^i(G^*) \quad \text{for} \quad i = 1, 3.
\end{aligned}
$$

Hence, $G^\emptyset \ll G^*$. Similarly, for all $G \in \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_6 \cup \Gamma_9 \cup \Gamma_{10}$, we can show that $G \ll G^*$.

Next, take an FTA network $H_0 \in \Gamma_7$. Again, we show that H_0 is indirectly dominated by G^*. To be specific, let us consider a sequence of FTA networks $\{H_r\}_{r=0}^4$ whose edge sets are given as follows:

$$
\begin{aligned}
E(H_0) &= \{(1,2), (2,3), (2,4)\}, \\
E(H_1) &= \{(2,3), (2,4)\}, \\
E(H_2) &= \{(2,3), (3,4), (4,2)\}, \\
E(H_3) &= \{(1,2), (2,3), (3,4), (4,2)\}, \\
E(H_4) &= \{(1,2), (2,3), (3,4), (4,2), (1,4)\}.
\end{aligned}
$$

By construction and from Table 1, we have

$$
\begin{aligned}
&H_0 \xrightarrow{\{1\}} H_1 \in \Gamma_4 \quad \text{and} \quad W^1(H_0) < W^1(G^*), \\
&H_1 \xrightarrow{\{3,4\}} H_2 \in \Gamma_5 \quad \text{and} \quad W^i(H_1) < W^i(G^*) \quad \text{for} \quad i = 3, 4, \\
&H_2 \xrightarrow{\{1,2\}} H_3 \in \Gamma_8 \quad \text{and} \quad W^i(H_2) < W^i(G^*) \quad \text{for} \quad i = 1, 2, \\
&H_3 \xrightarrow{\{1,4\}} H_4 \in \Gamma_{10} \quad \text{and} \quad W^i(H_3) < W^i(G^*) \quad \text{for} \quad i = 1, 4, \\
&H_4 \xrightarrow{\{1,3\}} G^* \in \Gamma_{11} \quad \text{and} \quad W^i(H_4) < W^i(G^*) \quad \text{for} \quad i = 1, 3.
\end{aligned}
$$

Hence, $H_0 \ll G^*$. Similarly, for all $H \in \Gamma_4 \cup \Gamma_5 \cup \Gamma_7 \cup \Gamma_8$, we can show that $H \ll G^*$. This completes the proof of part (ii). □

A.2 Proof of Theorem 2

First, we show that K_H coincides with the farsighted core. Lemma 1 implies that the complete network is included in the farsighted core. On the other hand, Lemma 1 implies that no FTA network other than G^* can be included in the farsighted core. Thus, $\{G^*\}$ is the farsighted core.

Next, we show that $\{G^*\}$ is a farsighted vNM stable set. Because K_H is a singleton, its internal stability is trivial. External stability directly follows from Lemma 1. Thus, K_H is a farsighted vNM stable set. Uniqueness directly follows from the coincidence of the farsighted core and the farsighted vNM stable set. □

A.3 Proof of Lemma 2

Before proceeding, we show some elementary facts. First, we have $n_i(J) = d_J(i) + 1$ for any graph $J \in \Gamma$ and for any $i \in N$. Then, Eq. (8) can be rewritten as follows: for any graph $J \in \Gamma$,

$$\sum_{i \in N} n_i(J) = n + 2e(J). \tag{18}$$

Next, given an arbitrary country (say, country k) and arbitrary graphs $H, G \in \Gamma$, let us define the following subsets A, B, C, D of N.

$A \equiv N_k(H) \cap N_k(G)$: the subset of countries that have FTAs with country k in both H and G.

$B \equiv N \setminus [N_k(H) \cup N_k(G)]$: the subset of countries that do not have FTAs with country k in both H and G.

$C \equiv N_k(H) \cap [N \setminus N_k(G)]$: the subset of countries that have FTAs in H, but do not in G.

$D \equiv N_k(G) \cap [N \setminus N_k(H)]$: the subset of countries that have FTAs in G, but do not in G.

Because $n_k(H) = |A \cup C| = |A| + |C|$ and $n_k(G) = |A \cup D| = |A| + |D|$, then we have $n_k(H) - n_k(G) = |C| - |D|$. We make use of the above facts in the following calculation.

Country k's welfare depends not only on the graph, but also on the common pre-agreement tariff rate t. Then, the difference $W^k(H) - W^k(G)$ multiplied by a positive constant $(n+1)^2$ can be considered a function of the common pre-agreement tariff rate t. Using Eq. (7), we obtain the following expression:

$$
\begin{aligned}
f(t) &\equiv (n+1)^2 \left[W^k(H) - W^k(G)\right] \\
&= \frac{1}{2}\left[n\alpha - (n - n_k(H))t\right]^2 - \frac{1}{2}\left[n\alpha - (n - n_k(G))t\right]^2 \\
&\quad + (n+1)(n - n_k(H))\left[\alpha - (1 + n_k(H))t\right] t \\
&\quad - (n+1)(n - n_k(G))\left[\alpha - (1 + n_k(G))t\right] t \\
&\quad + \sum_{i \in A} \left[\{\alpha + (n - n_i(H))t\}^2 - \{\alpha + (n - n_i(G))t\}^2\right] \\
&\quad + \sum_{i \in B} \left[\{\alpha - (1 + n_i(H))t\}^2 - \{\alpha - (1 + n_i(G))t\}^2\right] \\
&\quad + \sum_{i \in C} \left[\{\alpha + (n - n_i(H))t\}^2 - \{\alpha - (1 + n_i(G))t\}^2\right] \\
&\quad + \sum_{i \in D} \left[\{\alpha - (1 + n_i(H))t\}^2 - \{\alpha + (n - n_i(G))t\}^2\right].
\end{aligned}
$$

It is easy to verify that $f(0) = 0$. Therefore, by making use of the Taylor expansion of f around zero (up to the first order), we obtain $f(t) \doteqdot f'(0) \times t$. Consequently, for a sufficiently small positive t, we have $f(t) > 0$ if and only if $f'(0) > 0$. Simply calculating $f'(t)$, we obtain

$$
\begin{aligned}
& f'(t) \\
&= -\{n - n_k(H)\}\left[n\alpha - \{n - n_k(H)\}t\right] + \{n - n_k(G)\}\left[n\alpha - \{n - n_k(G)\}t\right] \\
&\quad + (n+1)\{n - n_k(H)\}\left[\alpha - \{1 + n_k(H)\}t\right] - (n+1)\{n - n_k(H)\}\{1 + n_k(H)\}t \\
&\quad - (n+1)\{n - n_k(G)\}\left[\alpha - \{1 + n_k(G)\}t\right] + (n+1)\{n - n_k(G)\}\{1 + n_k(G)\}t \\
&\quad + 2\sum_{i \in A} \langle \{n - n_i(H)\}\left[\alpha - \{n - n_i(H)\}t\right] - \{n - n_i(G)\}\left[\alpha + \{n - n_i(G)\}t\right] \rangle \\
&\quad + 2\sum_{i \in B} \langle -\{1 + n_i(H)\}\left[\alpha - \{1 + n_i(H)\}t\right] + \{1 + n_i(G)\}\left[\alpha - \{1 + n_i(G)\}t\right] \rangle \\
&\quad + 2\sum_{i \in C} \langle \{n - n_i(H)\}\left[\alpha + \{n - n_i(H)\}t\right] + \{1 + n_i(G)\}\left[\alpha - \{1 + n_i(G)\}t\right] \rangle \\
&\quad + 2\sum_{i \in D} \langle -\{1 + n_i(H)\}\left[\alpha - \{1 + n_i(H)\}t\right] - \{n - n_i(G)\}\left[\alpha + \{n - n_i(G)\}t\right] \rangle .
\end{aligned}
$$

By substituting $t = 0$ into $f'(t)$, we obtain

$$
f'(0)
$$

$$
\begin{aligned}
&= -\{n - n_k(H)\}\, n\alpha + \{n - n_k(G)\}\, n\alpha \\
&\quad + \{n - n_k(H)\}\,(n+1)\alpha - \{n - n_k(G)\}\,(n+1)\alpha \\
&\quad + 2\alpha \sum_{i\in A} \{n - n_i(H) - n + n_i(G)\} + 2\alpha \sum_{i\in B} \{-1 - n_i(H) + 1 + n_i(G)\} \\
&\quad + 2\alpha \sum_{i\in C} \{n - n_i(H) + 1 + n_i(G)\} + 2\alpha \sum_{i\in D} \{-1 - n_i(H) - n + n_i(G)\} \\
&= \alpha\,\{n - n_k(H)\}\,(-n + n + 1) + \alpha\,\{n - n_k(G)\}\,(n - n - 1) \\
&\quad + 2\alpha \sum_{i\in A} \{n_i(H) - n_i(G)\} + 2\alpha \sum_{i\in B} \{n_i(G) - n_i(H)\} \\
&\quad + 2\alpha \sum_{i\in C} \{n + 1 + n_i(G) - n_i(H)\} + 2\alpha \sum_{i\in D} \{n_i(G) - n_i(H) - (n+1)\} \\
&= \alpha\,\{n - n_k(H)\} - \alpha\,\{n - n_k(G)\} \\
&\quad + 2\alpha \left[\begin{array}{l}
\sum_{i\in A} n_i(G) - \sum_{i\in A} n_i(H) + \sum_{i\in B} n_i(G) - \sum_{i\in B} n_i(H) \\
+ \sum_{i\in C}(n+1) + \sum_{i\in C} n_i(G) - \sum_{i\in C} n_i(H) \\
- \sum_{i\in D}(n+1) + \sum_{i\in D} n_i(G) - \sum_{i\in D} n_i(H)
\end{array}\right] \\
&= \alpha\,\{n_k(G) - n_k(H)\} + 2\alpha \left[\sum_{i\in N} n_i(G) - \sum_{i\in N} n_i(H) + (n+1)\,\{|C| - |D|\}\right] \\
&= -\alpha\,\{n_k(H) - n_k(G)\} + 2\alpha \left[\sum_{i\in N} n_i(G) - \sum_{i\in N} n_i(H) + (n+1)\,\{n_k(H) - n_k(G)\}\right] \\
&= \alpha\,\{2(n+1) - 1\}\,\{n_k(H) - n_k(G)\} + 2\alpha \left\{\sum_{i\in N} n_i(G) - \sum_{i\in N} n_i(H)\right\} \\
&= \alpha(2n+1)\,\{n_k(H) - n_k(G)\} + 2\alpha \left\{\sum_{i\in N} n_i(G) - \sum_{i\in N} n_i(H)\right\} \\
&= \alpha(2n+1)\,\{n_k(H) - n_k(G)\} + 2\alpha\,\{n + 2e(G) - n - 2e(H)\} \\
&= \alpha(2n+1)\,\{n_k(H) - n_k(G)\} - 4\alpha\,\{e(H) - e(G)\} \\
&= \alpha(2n+1)\left[n_k(H) - n_k(G) - \frac{4}{2n+1}\,\{e(H) - e(G)\}\right] \\
&= \alpha(2n+1)\Delta.
\end{aligned}
$$

Therefore, we have $f'(0) > 0$ if and only if $\Delta > 0$. This proves Lemma 2-(i).

To prove Lemma 2-(ii), it suffices to show that $\sum_{k\in N} W^k(H) - \sum_{k\in N} W^k(G) > 0$

if and only if $\sum_{k\in N}\Delta > 0$.

$$\begin{aligned}
\sum_{k\in N}\Delta &= \sum_{k\in N}\left[\{n_k(H) - n_k(G)\} - \frac{4}{2n+1}\{e(H) - e(G)\}\right] \\
&= \sum_{k\in N} n_k(H) - \sum_{k\in N} n_k(G) - \sum_{k\in N}\frac{4}{2n+1}\{e(H) - e(G)\} \\
&= \{n + 2e(H)\} - \{n + 2e(G)\} - \frac{4n}{2n+1}\{e(H) - e(G)\} \\
&= 2\{e(H) - e(G)\} - \frac{4n}{2n+1}\{e(H) - e(G)\} \\
&= \left(2 - \frac{4n}{2n+1}\right)\{e(H) - e(G)\} \\
&= \frac{2}{2n+1}\{e(H) - e(G)\}.
\end{aligned}$$

This completes the proof. □

A.4 Proof of Observation 2

Part (i)

If country k forms new FTAs and if no other FTAs are formed, we have $n_k(H) - n_k(G) = e(H) - e(G) > 0$. By substituting this into Eq. (13), we obtain

$$\Delta = \left(1 - \frac{4}{2n+1}\right)\{n_k(H) - n_k(G)\} = \frac{2n-3}{2n+1}\{n_k(H) - n_k(G)\}. \tag{19}$$

Because $n \geqq 2$, we have $\Delta > 0$.

Part (ii)

If the number of FTAs not involving country k increases, we have $n_k(H) - n_k(G) = 0$ and $e(H) - e(G) > 0$. By substituting this into Eq. (13), we obtain

$$\Delta = -\frac{4}{2n+1}\{e(H) - e(G)\} < 0. \tag{20}$$

Part (iii)

For any G other than the complete network G^*, we have $e(G^*) > e(G)$. Then, Lemma 2 implies $\sum_{k\in N} W^k(G^*) > \sum_{k\in N} W^k(G)$. □

A.5 Proof of Theorem 3

Let G be a current network in which a pair of countries do not have a bilateral FTA between them. Then, by Observation 2-(ii), they have myopic incentives to form a new FTA between them. In other words, G can be directly dominated by another graph.

Let us consider the complete network G^*. Then, no pair of countries can form a bilateral FTA; moreover, by Observation 2 again, no country has an incentive to abandon the existing FTAs. The complete network is not directly dominated. □

A.6 Proof of Theorem 5

As mentioned in the text, any graph cannot indirectly dominate other graphs in the same isomorphic class. Therefore, internal stability is achieved.

Let us turn to external stability. As shown in the proof of Lemma 3, any graph $G \in \Gamma_6 \cup \Gamma_8 \cup \Gamma_9 \cup \Gamma_{10} \cup \Gamma_{11}$ is indirectly dominated by a graph in $K_L = \Gamma_5$. It remains to be shown that for any $G \in \Gamma_1 \cup \Gamma_2 \cup \Gamma_3 \cup \Gamma_4 \cup \Gamma_7$, there exists a graph $H \in K_L$ that indirectly dominates G.

Take the empty graph $G^{\emptyset} \in \Gamma_1$ and consider the following sequence of graphs:

$$\begin{aligned}
&G_1 \in \Gamma_2 : E(G_1) = \{(1,2)\}, && e(G_1) = 1,\\
&G_2 \in \Gamma_4 : E(G_2) = \{(1,2),(2,3)\}, && e(G_2) = 2,\\
&H \in \Gamma_5 : \ E(H) = \{(1,2),(1,3),(2,3)\}, && e(H) = 3.
\end{aligned}$$

Similar to the proof of Lemma 3, we can show that

$$\begin{aligned}
&G^{\emptyset} \xrightarrow{\{1,2\}} G_1 \quad \text{and} \quad W^i(G^{\emptyset}) < W^i(H) \quad \text{for} \quad i = 1,2,\\
&G_1 \xrightarrow{\{2,3\}} G_2 \quad \text{and} \quad W^i(G_1) < W^i(H) \quad \text{for} \quad i = 2,3,\\
&G_2 \xrightarrow{\{1,3\}} H \quad \text{and} \quad W^i(G_2) < W^i(H) \quad \text{for} \quad i = 1,3.
\end{aligned}$$

Hence, $G^{\emptyset}$ is indirectly dominated by $H \in K_L$. The above argument can be modified to show that any graph in $G \in \Gamma_1 \cup \Gamma_2 \cup \Gamma_4$ is indirectly dominated by a graph in K_L.

Next, let us consider $J_0 \in \Gamma_7$ and a sequence of graphs such that

$$\begin{aligned}
&J_0 \in \Gamma_7 : E(J_0) = \{(1,3),(2,3),(4,3)\}, && e(J_0) = 3,\\
&J_1 \in \Gamma_3 : E(J_1) = \{(1,2),(3,4)\}, && e(J_1) = 2,\\
&J_2 \in \Gamma_4 : E(J_2) = \{(1,2),(2,3)\}, && e(J_2) = 2,\\
&H \in \Gamma_5 : E(H) = \{(1,2),(1,3),(2,3)\}, && e(H) = 3.
\end{aligned}$$

Again, similar to the proof of Lemma 3, we can show that

$$\begin{aligned}
&J_0 \xrightarrow{\{1,2\}} J_1 \quad \text{and} \quad W^i(J_0) < W^i(H) \quad \text{for} \quad i = 1,2,\\
&J_1 \xrightarrow{\{2,3\}} J_2 \quad \text{and} \quad W^i(J_1) < W^i(H) \quad \text{for} \quad i = 2,3,\\
&J_2 \xrightarrow{\{1,3\}} H \quad \text{and} \quad W^i(J_2) < W^i(H) \quad \text{for} \quad i = 1,3.
\end{aligned}$$

Hence, J_0 is indirectly dominated by $H \in K_L$. The above argument also implies that any graph in $\Gamma_3 \cup \Gamma_7$ is indirectly dominated by a graph in K_L. Consequently, any graph not in Γ_5 is indirectly dominated by a graph in Γ_5. External stability is achieved. □

References

Aghion, P., P. Antràs, and E. Helpman, 2007, Negotiating free trade, *Journal of International Economics* 73: 1–30.

Bagwell, K. and R.W. Staiger, 1998, Regionalism and multilateral tariff cooperation, in: J. Piggott and A. Woodland (eds.), *International Trade Policy and the Pacific Rim*, Macmillan, London.

Bhagwati, J.N., 1991, *The World Trading System at Risk*, Princeton University Press, Princeton.

Bhagwati, J.N., 1993, Regionalism and multilateralism: an overview, in: J. de Melo and A. Panagariya (eds.), *New Dimensions in Regional Integration*, Cambridge University Press, Cambridge, UK.

Bollobás, B., 1979, *Graph Theory: An Introductory Course*, Springer-Verlag, Berlin.

Bond, E., R. Riezman, and C. Syropoulos, 2004, A strategic and welfare theoretic analysis of free trade areas, *Journal of International Economics* 64: 1–27.

Das, S.P. and S. Ghosh, 2006, Endogenous trading bloc formation in a North-South global economy, *Canadian Journal of Economics* 39: 809–830.

Furusawa, T. and H. Konishi, 2005, Free trade networks with transfers, *Japanese Economic Review* 56: 144–164.

Furusawa, T. and H. Konishi, 2007, Free trade networks, *Journal of International Economics* 72: 310–335.

Goyal, S. and S. Joshi, 2006, Bilateralism and free trade, *International Economic Review* 47: 749–778.

Greenberg, J., 1990, *The Theory of Social Situations—An Alternative Game-Theoretic Approach*, Cambridge University Press, Cambridge, UK.

Herings, P.J-J, A. Mauleon, and V. Vannetelbosch, 2009, Farsightedly stable networks, *Games and Economic Behavior* 67: 526–541.

Jackson, M. and A. Wolinsky, 1996, A strategic model of social and economic networks, *Journal of Economic Theory* 71: 44–74.

Krugman, P., 1993, Regionalism versus multilateralism: analytical notes, in: J. de Melo and A. Panagariya (eds.), *New Dimensions in Regional Integration*, Cambridge University Press, Cambridge, UK.

Macho-Stadler, I. and L. Xue, 2007, Winners and losers from the gradual formation of trading blocs, *Economica* 74: 664–681.

Oladi, R. and H. Beladi, 2008, Is regionalism viable? A case for global free trade, *Review of International Economics* 16: 293–300.

Seidmann, D.J., 2009, Preferential trading arrangements as strategic positioning, *Journal of International Economics* 79: 143–159.

Saggi, K., 2006, Preferential trade agreements and multilateral tariff cooperation, *International Economic Review* 47: 29–57.

Saggi, K. and H.M. Yildiz, 2010, Bilateralism, multilateralism, and the quest for global free trade, *Journal of International Economics* 81: 26–37.

Viner, J., 1950, *The Customs Union Issue*, Carnegie Endowment for International Peace, New York.

von Neumann, J. and O. Morgenstern, 1944, *Theory of Games and Economic Behavior* (2nd edition: 1947, 3rd edition: 1953), Princeton University Press, Princeton.

WTO, 2010, `http://www.wto.org/english/tratop_e/region_e/region_e.htm` (as of October 15, 2010).

Xue, L. and J. Zhang, 2009, Farsighted free trade network, unpublished manuscript.

Yi, S-S, 1996, Endogenous formation of customs unions under imperfect competition: open regionalism is good, *Journal of International Economics* 41: 153–177.

Yi, S-S, 2000, Free-trade areas and welfare: an equilibrium analysis, *Review of International Economics* 8: 336–347.

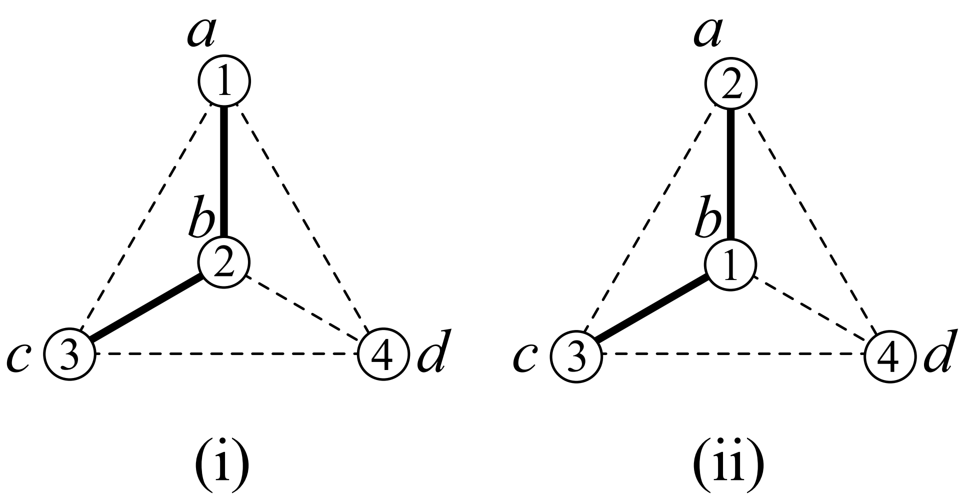

Figure 1: Examples of FTA networks

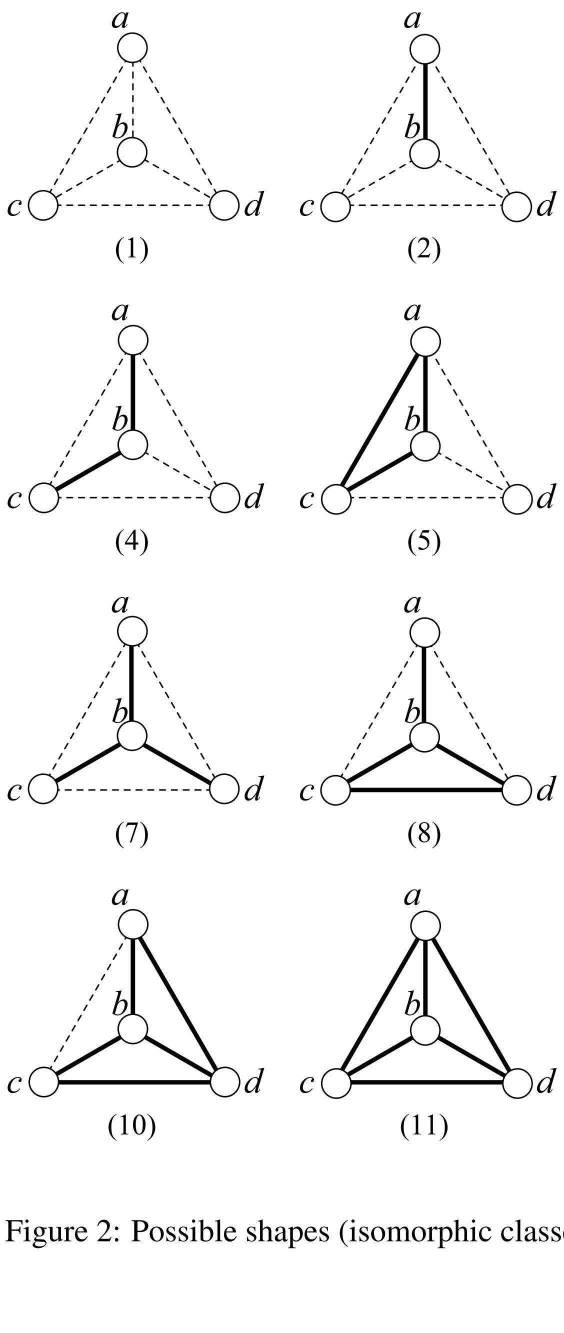

Figure 2: Possible shapes (isomorphic classes) of FTA networks

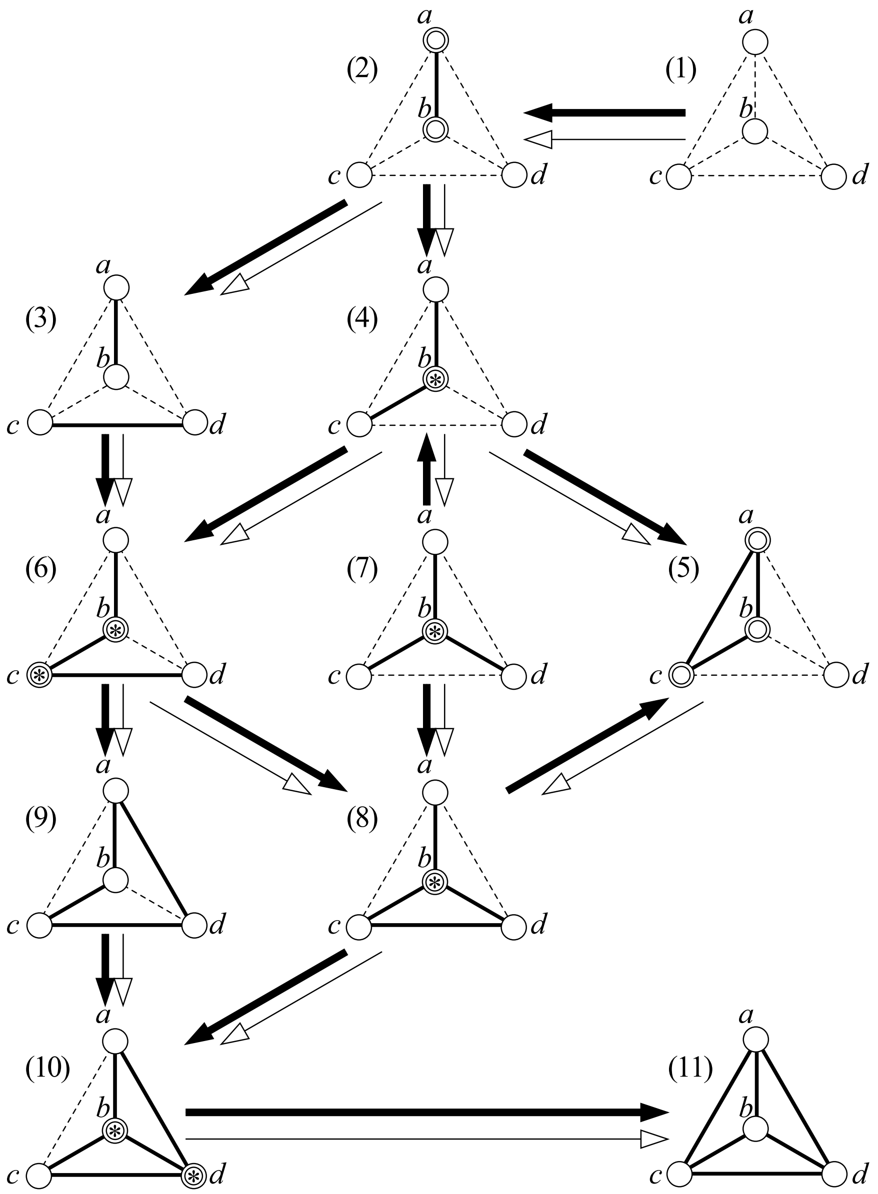

Figure 3: Direct domination relation among FTA networks. Note: (i) Bold black arrows indicate the direct domination relation when the pre-agreement tariffs are high; (ii) Thin arrows with white heads indicate the direct domination relation when the pre-agreement tariffs are low; (iii) Asterisks mean that the welfare of countries located at these addresses is higher than under global free trade in the case of high pre-agreement tariffs; (iv) Double circles mean that the welfare of countries located at these addresses is higher than under global free trade in the case of low pre-agreement tariffs.

	Addresses in each isomorphic class				
	a	b	c	d	World welfare
Γ_1	3/8 (0.375)	3/8 (0.375)	3/8 (0.375)	3/8 (0.375)	3/2 (1.5)
Γ_2	4/9 (0.444)	4/9 (0.444)	3/8 (0.375)	3/8 (0.375)	59/36 (1.638)
Γ_3	4/9 (0.444)	4/9 (0.444)	4/9 (0.444)	4/9 (0.444)	16/9 (1.777)
Γ_4	19/48 (0.395)	∗163/288 (∗0.565)	19/48 (0.395)	3/8 (0.375)	499/288 (1.732)
Γ_5	15/32 (0.468)	15/32 (0.468)	15/32 (0.468)	3/8 (0.375)	57/32 (1.781)
Γ_6	19/48 (0.395)	∗149/288 (∗0.517)	∗149/288 (∗0.517)	19/48 (0.395)	263/144 (1.826)
Γ_7	28/75 (0.373)	∗52/75 (∗0.693)	28/75 (0.373)	28/75 (0.373)	136/75 (1.813)
Γ_8	28/75 (0.373)	∗1073/1800 (∗0.596)	357/800 (0.446)	357/800 (0.446)	6703/3600 (1.861)
Γ_9	15/32 (0.468)	15/32 (0.468)	15/32 (0.468)	15/32 (0.468)	15/8 (1.875)
Γ_{10}	21/40 (0.423)	∗339/800 (∗0.525)	21/40 (0.423)	∗339/800 (∗0.525)	759/400 (1.8975)
Γ_{11}	12/25 (0.480)	12/25 (0.480)	12/25 (0.480)	12/25 (0.480)	48/25 (1.920)

Table 1: Welfare levels when the pre-agreement tariffs are high. Note: (i) The fractions are the exact numbers, while the decimals in parentheses are approximate numbers; (ii) To obtain $W^k(G)$, the values in the table must be multiplied by α^2; (iii) The values with an asterisk (∗) are higher than the corresponding values obtained in the complete graph; (iv) World welfare is the sum of the welfare of four countries.