WHY ARE TRADE AGREEMENTS REGIONAL?

by

Ben Zissimos

Working Paper No. 10-W02

March 2010

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

Why are Trade Agreements Regional?[1]

Ben Zissimos[2]

Vanderbilt University

March 2010

Abstract: This paper shows how distance may be used to coordinate on a unique equilibrium in which trade agreements are regional. Trade agreement formation is modeled as coalition formation. In a standard trade model with no distance between countries a familiar problem of coordination failure occurs, giving rise to multiple equilibria; any one of many possible trade agreements can form. With distance between countries, regional trade agreements generate larger rent-shifting effects than non-regional agreements. Countries use these effects to coordinate on a unique equilibrium.

Keywords: Coalition, coordination, regionalism, preferential trade agreement, trade liberalization.

JEL Classification Numbers: F02, F13, F15, C72.

[1]Thanks are due to many friends and colleagues but especially to Rick Bond and three anonymous referees for detailed comments on earlier drafts. I am also grateful for comments from seminar participants of the International Industrial Organization Conference, Georgia Tech, the Midwest International Economics Meetings, Vanderbilt, the CTN Meetings at CORE, Louvain la Neuve, and departmental seminars at Southern Illinois University, Syracuse University, University of Notre Dame and University of Oregon. This paper was revised while I was visiting Fundação Getulio Vargas, Rio de Janeiro, who provided a conducive research environment. Financial support by the Center for the Americas at Vanderbilt University is gratefully acknowledged.

[2]Dept. of Economics, Vanderbilt University, Nashville, TN 37235.
Tel: ++1 615 322 3339.
E-mail: ben.zissimos@vanderbilt.edu

1. Introduction

Recent empirical research presents substantive evidence that trade agreements (TAs) are regional.[3] The theoretical literature provides support for this finding, showing that the potential gains to a regional agreement are higher than to a non-regional agreement, and that non-regional agreements may even be trade diverting. But no attempt has been made before to provide a theory of how TAs might actually form into regional structures.[4]

This paper puts forward a theory of regional TA formation. It argues that there is a coordination problem at the heart of the TA formation process, and countries seek TAs that are regional as a way to solve that problem. There is undoubtedly significant 'pre-play communication' between policy-makers before a TA is formed. This observation is used in the past literature to set aside problems of coordination. But in fact, the need for pre-play communication actually implies that there is a coordination problem to be resolved as part of the TA formation process. The main point brought to light in this paper, by setting the issue of coordination center stage, is that countries can use geographical organization to solve their coordination problem. Thus, each country seeks other countries in its region, and only countries in its region, when forming a TA.

The model is based on Brander and Spencer (1984) and Yi (1996). Brander and Spencer

[3] Prominent examples of regional TAs are the European Union (EU), the Mercado Comun del Cono Sur (MERCOSUR) and the North American Free Trade Agreement (NAFTA). More generally, for a sample of 54 countries, Baier and Bergstrand (2004) show that (the inverse of) distance is a good predictor of TA membership. Non-regional agreements do exist of course, such as the recently signed TA between the US and South Korea. But evidently they are the exception rather than the rule.

TA is a catch-all term that refers to all agreements in which a group of countries commit to trade among members preferentially. This encompasses free trade agreements (FTAs) in which members agree to remove internal tariff barriers but set external tariffs independently, and customs unions (CUs) which are like FTAs but with the additional requirement that members coordinate on common external tariffs. In practice, FTAs are more common but most of the academic literature focuses on CUs because they are analytically easier to handle. To focus the discussion on the regional nature of these agreements rather than the technical details of their operation, we will use the catch-all term TA wherever possible (but distinctions between trading arrangements will be made where relevant.)

[4] There is a literature that looks at the feasibility of preferential trade agreements when countries cannot write binding contracts over tariffs; see for example Bagwell and Staiger, (1997a,b), (1999), Bond and Syropoulos (1996), Bond (2001), Bond, Syropoulos and Winters (2001), Bond, Riezman and Syropoulos (2004) and Saggi (2006). These previous papers all look at how agreements between sufficiently patient countries may be sustained through repeated interactions in the face of a short-run incentive to deviate. In the model of this present paper, there is no short-run incentive to deviate. The problem focused on instead is whether a country is able to form an agreement with the other countries that it would like to have as members - the problem of coordination. This is different from the coordination failure considered by McClaren (2002), which is between the firms and their government in each country.

(1984) show, in a two-country model, that rents made by foreign firms in the domestic market can be shifted back home by the government using tariffs. Yi (1996) uses a Brander-Spencer type model to show that a group of countries may obtain a higher payoff from TA formation than from moving to free trade. The present paper takes a special case of Yi's model in which goods are homogeneous and extends it by putting it in a regional setting.[5]

One of Yi's key results shows that a country would always prefer to leave its own TA in order to join another TA of equal or larger size, since the new TA that forms eliminates greater harmful rent-shifting effects and confers greater terms-of-trade benefits. However, in the present paper, a new effect is revealed when a regional dimension is introduced to the model. Without an agreement, since more rents are dissipated through transportation between regions than within them, there is more scope for rent-shifting within a region than across regions. TA formation within a region eliminates this greater harmful rent shifting among members, and in addition has greater beneficial terms-of-trade effects. Therefore, the value to a member of joining a regional TA of a given size is greater than the value of a TA across regions. This effect tends to push the countries of a region towards the formation of a regional TA.

In order to see the intuition behind this effect, consider the original proposals made in the 1960s for NAFTA - the North *Atlantic* Free Trade Agreement - between Canada, the UK and the US. Interpreted within the context of the present model, Canada and the US would have liked the UK to form a TA with them, but the UK ultimately obtained a higher payoff from the formation of an agreement with nearby European nations. This was so because the gains to elimination of rent shifting within Europe and the terms-of-trade gains over North America were of greater value to the UK.[6]

To introduce the problem of coordination failure in the present context, TA formation is

[5] Yi (1996) compares how 'open regionalism' can help with the attainment of free trade compared to the outcome under 'exclusive regionalism' in which TA membership must be unanimous. The present paper draws on Yi's analysis of exclusive regionalism and it does not address the question of whether open regionalism would be beneficial in a regional setting. In his study of exclusive regionalism, Yi (1996) identifies the stable equilibrium structure of TAs; an approach pioneered by Riezman (1985) that will be extended to a regional setting in the present paper.

[6] The underlying intuition is robust to the fact that the NAFTA proposals were obviously for an FTA while the EU (formerly the European Economic Community, or EEC) is a CU. In a broader setting, the choice of trading arrangement may have a significant bearing on the outcome. This point is made by Riezman (1999), who endogenizes the decision by countries over whether to adopt a CU or FTA, showing that the choice of regime may affect whether free trade can be reached. (Also see Bloch's 2003 discussion of CUs versus FTAs, and Bond, Riezman and Syropoulos 2004.)

modeled based on Hart and Kurz's (1983) *simultaneous move exclusive membership game*. In their original game, simultaneously and without communicating, each player writes down a list of other players with whom she would like to form a coalition. The lists form intersecting sets of players and each of the intersecting sets forms a coalition. But if two players fail to name each other then neither ends up in the same coalition even if it would have been mutually beneficial.[7]

In the model of the present paper each and every country, simultaneously and without communicating, writes down a list of others with whom it would like to form a TA. When transport costs between all countries are zero, so in effect there is no regional dimension to the model, the problem of coordination failure arises between them. Any one of many possible TAs may arise in equilibrium. When transport costs of trading between regions are greater than zero (but not large enough to prohibit trade between regions) countries use the difference in rent-shifting effects within and between regions to coordinate on regional TA formation. TAs form simultaneously, one in each region, and each TA includes all countries in that region. This is the sense in which the coordination problem is resolved when a regional dimension is introduced to the model. The model is highly stylized, particularly in terms of its regional structure. Nevertheless, even though strong assumptions are made about functional forms and the TA formation process, the results seem intuitively plausible and may be indicative of a general driving force towards regionalism for which there appears to be substantial evidence.

In general terms, the literature on regionalism addresses two issues. The first issue, which was the focus of Viner (1950) in his seminal work on the topic, concerns the welfare effects of TA formation and expansion. The second issue is with the stability of TAs. Given endogenous TA formation, what TA structures are stable? Are trade blocks conducive to or inimical to the eventual attainment of free trade? (See Bhagwati 1993, although the roots of this question are found in Viner 1950). We will address both issues in this paper.

Firstly, standard results will be shown to carry over to the regional setting of the present paper in that TA formation and expansion tends to increase aggregate member welfare and hurts non-members. But we will add that a regional TA is worth more to its members than a non-regional block through rent-shifting and terms-of-trade effects.

[7] See Bloch 2003 and Yi 2003 for reviews of the literature on coalition formation.

Secondly, we will examine the issue of stability. Yi (1996) argues that an equilibrium TA structure must be asymmetric. Countries use the advantage in the sequence of TA formation that they are exogenously granted to form a larger TA. The countries in the larger TA are better off even than under free trade because they enjoy more favorable terms-of-trade effects over non-member countries. In the present paper, no such advantages arise due to the fact that TA formation is simultaneous and so each country is uncertain about the outcome of the TA formation process. As a result TA formation can be symmetric, with no larger TA arising that would prefer the status quo to free trade. Thus the door is left open to the consolidation of free trade at a later (unmodeled) stage.[8]

The paper proceeds as follows. Section 2 introduces the basic trade model and uses it to explore the economic effects of TA formation in regions. Section 3 introduces the TA formation game. Section 4 shows that, in the TA formation game, when transport costs are zero there are multiple equilibria and no predictions can be made as to which TA will prevail. Section 5 then shows that when transport costs are greater than zero this provides a coordination device through which countries are able to coordinate on a unique equilibrium in which regional TAs form. Conclusions are drawn in Section 6.

2. A Model of Trade Agreements in Regions

We will extend a familiar model of TA formation based on Cournot competition. Let N be the set of countries. Each country, i, has a representative consumer, firm, and government, each denoted by its corresponding country identifier as $i \in N$.

There are six countries; $N = \{1, 2, 3, 4, 5, 6\}$. This is different from a standard TA formation model, which would typically have just three countries. In our model, there is a *regional structure* that partitions our set, N, of six countries into two regions; $R_1 = \{1, 2, 3\}$

[8]The framework of the present paper could potentially be extended to examine the possibility that TA formation gives way to world free trade. In addition to Riezman (1999), see Aghion, Antras and Helpman (2007) and Seidmann (2009) for contributions that model the dynamics of regionalism. Building on Baldwin (1996), Krishna (1998) shows how political interests can undermine the progression from regionalism. Ornelas (2005a,b) shows that TAs may create problems for multilateral trade liberalization 'through their own success;' if governments can adjust tariffs then they only support trade-creating TAs, but then non-member countries may prefer to free-ride on such agreements, blocking a subsequent move to free trade. Ethier (1998) considers how multilateral liberalization may give way to regionalism. See Bagwell and Staiger (1998) on how TAs undermine the principles by which multilateral trade liberalization is achieved. Also, see Bagwati, Greenaway and Panagariya (1998) for a literature review on the dynamics of regionalism. See Baldwin (2009) for a comprehensive review of the literature.

and $R_2 = \{4, 5, 6\}$. A three-country framework is the simplest possible framework in which TA formation can be examined, since a minimum of two countries are required to form a TA and at least one country must remain outside so that the effects on a non-member can be analyzed. To extend this simple basic approach to a regional setting requires a set-up based on two regions, each of which has three countries.[9]

Regions are some distance apart from one another. Let d_{ij} measure the distance between any two countries $i, j \in N$. To keep the analysis as simple as possible, we will say that if countries i and j are *not* in the same region then $d_{ij} = d$ while if i and j are in the same region then $d_{ij} = 0$.

The details of the TA formation game will be defined fully in due course, but it may be helpful to preview the game's extensive form to put the model in context. First, TA formation takes place. Next, taking trading arrangements as given, firms make production decisions. Finally, consumption takes place. We will adopt the usual inductive approach of solving this sequence backwards. Thus, in what follows it will make sense to assume that firms take the structure of trade agreements, tariffs and demand curves as given.

2.1. Preferences and Production

There are two goods in the model, denoted M and X. Good M is chosen as the numeraire. Countries are endowed with equal quantities of M, which is transferred internationally to settle the balance of trade. The term M_i measures consumption of M in country i. By assumption, each country is endowed with a sufficient quantity of M to ensure that it consumes a positive quantity in equilibrium.[10]

All the firms in the model, one in each country, produce the homogeneous product X. We will use x_{ij} to denote the quantity produced for the market in country i by the firm in

[9] This framework is general enough to demonstrate regionalism while being simple enough to yield clear-cut analytical solutions. In the concluding section, we will discuss how the forces for regionalism under discussion may be examined in a more realistic regional setting.

[10] Note that since all countries are endowed with M and produce X, there is no scope in the present model for trade diversion. That is, TA formation cannot lower welfare by inducing countries to import more from TA partners that do not have a comparative advantage. The gains and losses to TA formation here are driven instead by strategic considerations; this is a common feature of the recent literature. In the conclusions we will discuss possible extensions to a Heckscher-Ohlin setting in which trade diversion is possible.

country j, and X_i as the quantity produced by all firms for sale in country i:

$$X_i = \sum_{j \in N} x_{ij}. \tag{2.1}$$

Consumer preferences are approximated by the following quasi-linear function:

$$u_i = v(X_i) + M_i = eX_i - \frac{1}{2}X_i^2 + M_i, \tag{2.2}$$

where e is a parameter. This functional form is relatively simple, focusing attention on the impact of product differentiation by distance.[11]

The inverse demand curve of consumer i is obtained in the usual way by differentiating (2.2) with respect to x_{ij}:

$$p_i\left(X_i\right) = \frac{dv}{dx_{ij}} = e - X_i. \tag{2.3}$$

Firm j's (marginal) cost to produce a unit of X for sale in country i consists of three components: a private per unit cost, c, which is the same for all firms; the tariff, t_{ij}, levied by government i on imports from j; the transport cost, d_{ij}, of shipping from j to i. Thus, firm j's per-unit production cost for each market i is given by the function

$$c_{ij} = c + t_{ij} + d_{ij}. \tag{2.4}$$

We will assume that firms perceive markets as being segmented, and so they compete by choosing quantities in each country.[12] Firm j chooses x_{ij} to maximize profits in each market i, denoted π_{ij}:

$$\underset{\{x_{ij}\}}{\text{Max}}\ \pi_{ij} = \left(p_i - c_{ij}\right) x_{ij}, \tag{2.5}$$

where p_i is determined according to the inverse demand curve $p_i\left(X_i\right)$ given by (2.3).

Setting the first derivative of (2.5) equal to zero obtains the first order condition for firm j; $p_i - c_{ij} - x_{ij} = 0$. Summing first order conditions over all $j \in N$, in Cournot equilibrium,

$$x_{ij} = \frac{(e-c) + \sum_{k \in N} d_{ik} + \sum_{k \in N} t_{ik}}{7} - d_{ij} - t_{ij}. \tag{2.6}$$

[11] This basic functional form for preferences is used quite widely in the literature on TAs. Yi (1996) has a more general form which allows X to be horizontally differentiated. The model of this present paper could be extended in that direction but this would complicate the analysis considerably and would risk obscuring the effects resulting from the organization of countries into regions.

[12] This assumption is made for analytical simplicity, but approximates the weaker assumption that firms compete over capacities.

Output for market i by firm j depends negatively on d_{ij} and t_{ij}; the smaller the distance to market, and the lower the tariff, the larger the rents available from shipping to country i and so the higher the quantity produced. In contrast, output by firm j depends positively on the distance from country i to all other markets and the tariff set by country i on imports from all countries other than j. Note that the strength of demand relative to cost helps to determine the rents available to firm j as well; $e - c$ is common to all markets and can be made large enough to ensure that $x_{ij} > 0$ for all i, j.[13]

2.2. Trade Agreements and Trade Volumes

The structure of TAs in the world economy is defined as follows. A *TA structure* $B = (B_1, B_2, ..., B_m)$ is a partition of the set of countries N, where B_1, B_2, ..., B_m are TAs; $B_i \cap B_j = \emptyset$ for $i \neq j$, and $\cup_{i=1}^{m} B_i = N$. If B_i has only one element then it is referred to as a *singleton*; a country that does not coordinate trade policy with others.[14]

Recall that the location of each country is fixed either in R_1 or in R_2. Therefore, $(B_k \cap R_1) \cup (B_k \cap R_2) = B_k$. Let b_{ir} be the number of country i's TA partners that are in the same region as country i, and let b_{inr} be the number of country i's TA partners that are in the "other" region.[15] In the present simple regional set-up, $b_{ir} \in \{1, 2, 3\}$ and $b_{inr} \in \{0, 1, 2, 3\}$.

Using (2.6), we can now express outputs produced for country i in terms of regional and TA relationships. Let r stand for regional and nr stands for non-regional. Then t_{ir} is the tariff that country i sets on imports from non-members in the same region and t_{inr} is the tariff set on imports from non-members in the other region. Let m stand for TA member and let nm stand for non-member. Then we may use these mnemonics to classify outputs into four basic terms.

[13]The solution for x_{ij} obviously depends on the assumption that there is only one firm in each country. Some work has looked at how TA formation is affected by a change in the number of firms; see in particular Krishna (1998). From this earlier work, variation in firm numbers is most interesting when it is asymmetric. But since in the present paper we have already introduced a regional asymmetry to the model, we will leave aside formal analysis of variation in the number of firms across countries.

[14]In coalition formation, relations between countries are transitive; if countries 1 and 2 have an agreement and 2 and 3 have an agreement then 1 and 3 must have an agreement. In network formation, by contrast, relations may be intransitive; even if countries 1 and 2 have an agreement, it does not follow that 1 and 3 must have an agreement. Because the TA formation that we will consider involves coordination over external and internal tariffs, it implies a transitive relationship between members.

[15]Formally, if $i \in B_k$ and $i \in R_l$ then let b_{ir} be the cardinality of the set $B_k \cap R_l$ and let b_{inr} be the cardinality of the set $B_k \cap R_m$, $l \neq m$.

Write x_{irm} for output produced for country i by a country that is in the same region as country i *and* is a member of country i's TA:

$$x_{irm}\left(t_{ir},t_{inr};d\right)=\frac{(e-c)+3d+\left(3-b_{ir}\right)t_{ir}+\left(3-b_{inr}\right)t_{inr}}{7}. \tag{2.7}$$

Write x_{inrm} for output produced for country i by a country not in the same region but which is a member of country i's TA:

$$x_{inrm}\left(t_{ir},t_{inr};d\right)=\frac{(e-c)-4d+\left(3-b_{ir}\right)t_{ir}+\left(3-b_{inr}\right)t_{inr}}{7}. \tag{2.8}$$

Write x_{irnm} for output produced for country i by a country that is in the same region but not a member of country i's TA:

$$x_{irnm}\left(t_{ir},t_{inr};d\right)=\frac{(e-c)+3d-\left(4+b_{ir}\right)t_{ir}+\left(3-b_{inr}\right)t_{inr}}{7}. \tag{2.9}$$

Finally, write x_{inrnm} for output produced for country i by a country that is not in the same region and is not a member of country i's TA:

$$x_{inrnm}\left(t_{ir},t_{inr};d\right)=\frac{(e-c)-4d+\left(3-b_{ir}\right)t_{ir}-\left(4+b_{inr}\right)t_{inr}}{7}. \tag{2.10}$$

Total output is given by

$$\begin{aligned}X_i\left(t_{ir},t_{inr};d\right) &= b_{ir}x_{irm}\left(t_{ir},t_{inr};d\right)+\left(3-b_{ir}\right)x_{irnm}\left(t_{ir},t_{inr};d\right)\\ &\quad +b_{inr}x_{inrm}\left(t_{ir},t_{inr};d\right)+(3-b_{inr})x_{inrnm}\left(t_{ir},t_{inr};d\right).\end{aligned}$$

By (2.7), the greater is d and the higher are t_{ir} and t_{inr} the greater the output produced by a regional member of i's TA for country i. (This expression also describes output by firm i for its own national market.) By (2.8), the greater is d the smaller is the output produced by a non-regional member of i's TA for country i. Expressions (2.9) and (2.10) reflect the same basic intuition.

2.3. Welfare

Profits of domestic firms and tariff revenues are rebated back to consumers. Also, there is perfect competition in the world market for transportation. Based on these assumptions and the model set-up, country i's welfare can be expressed in terms of four economic components:

domestic consumer surplus, C_i; tariff revenue, T_i; shipping revenue, D_i; the domestic firm's profit at home and abroad, π_{ii} and $\sum_{j\in N\setminus i}\pi_{ji}$ respectively ($j \neq i$)). Country i's welfare is denoted w_i:

$$w_i = C_i + T_i + D_i + \pi_{ii} + \sum_{j\in N\setminus i}\pi_{ji}, \tag{2.11}$$

where $C_i = \frac{1}{2}(e - p_i) X_i$, $T_i = \sum_{j\in N} t_{ij}x_{ij}$ and $D_i = \sum_{j\in N} d_{ij}x_{ij}$. Because the transport sector is perfectly competitive, goods are delivered at cost and there is no surplus associated with that sector; $D_i = 0$. This specification makes 'iceberg' transportation costs consistent with the present general equilibrium setting.

The next result shows that (2.11) may be expressed strictly in terms of outputs.

Lemma 1. *Country i's welfare, w_i, may be written as*

$$w_i = v(X_i) - cX_i - \sum_{j\in N} d_{ij}x_{ij} - \sum_{j\in N\setminus\{i\}} (x_{ij})^2 + \sum_{j\in N\setminus\{i\}} (x_{ji})^2. \tag{2.12}$$

This result is familiar from the past literature. Equation (2.12) incorporates transport costs in an otherwise standard expression. Using (2.7)-(2.10), we can now express w_i as a function of tariffs and the regional structure of the model.

2.3.1. Optimal tariffs

The members of a TA coordinate on setting external tariffs. The problem of the representative TA member, country i, may be expressed as follows:

$$\underset{\{t_{ij}\}_{i\in B_k, j\notin B_k}}{\text{Max}} \sum_{i\in B_k} w_i = \sum_{i\in B_k}\left(v(X_i) - cX_i - \sum_{j\in N} d_{ij}x_{ij} - \sum_{j\in N\setminus\{i\}} (x_{ij})^2 + \sum_{j\in N\setminus\{i\}} (x_{ji})^2\right), \tag{2.13}$$

where $t_{ij} = 0$ for all $i, j \in B_k$. Using (2.13) we are now in a position to determine optimal tariffs.

Proposition 1. *Assume that country i belongs to a TA of b_{ir} regional members and b_{inr} non-regional members. Country i's unique optimal external tariff on imports from a non-member in the same region as country i is*

$$t_{ir}^{*}\left(b_{ir},b_{inr};d\right)=\frac{\left(1+2\left(b_{ir}+b_{inr}\right)\right)\left(e-c\right)}{\Delta\left(b_{ir},b_{inr}\right)}+\frac{3+6b_{ir}+b_{inr}\left(2\left(b_{ir}+b_{inr}\right)-7\right)}{2\Delta\left(b_{ir},b_{inr}\right)}d,$$

where $\Delta\left(b_{ir},b_{inr}\right)\equiv 7+\left(1+\left(b_{ir}+b_{inr}\right)\right)\left(1+2\left(b_{ir}+b_{inr}\right)\right).$

The unique optimal external tariff imposed by country i on imports from a non-member who is not in the same region as country i is

$$t_{inr}^{*}\left(b_{ir},b_{inr};d\right)=\frac{\left(1+2\left(b_{ir}+b_{inr}\right)\right)\left(e-c\right)}{\Delta\left(b_{ir},b_{inr}\right)}-\frac{5+b_{ir}\left(2b_{ir}-3\right)+2b_{inr}\left(5+b_{ir}\right)}{2\Delta\left(b_{ir},b_{inr}\right)}d.$$

Note that if $d=0$ then $t_{ir}^{*}\left(b_{ir},b_{inr};d\right)=t_{inr}^{*}\left(b_{ir},b_{inr};d\right)$. On the other hand, if $d>0$ then $t_{ir}^{*}\left(b_{ir},b_{inr};d\right)-t_{inr}^{*}\left(b_{ir},b_{inr};d\right)$ is increasing in $d>0$. Also notice that if $d=0$ then $t_{ir}^{*}=t_{inr}^{*}$ corresponds exactly to the optimal tariff found in previous literature.[16]

If countries are identical, as in Yi (1996), then the solution to (2.13) mandates that all members of a TA set the same external (joint-welfare-maximizing) tariff; they form a CU in other words. Here in the present setting, when countries are not identical (i.e. when $d>0$), members have an incentive to set a tariff that discriminates between non-members based on their location. With equal tariffs, a firm would always export a larger volume to a nearby country because less of its rents are dissipated in shipping; its export supply elasticity is increasing in distance. This in turn motivates higher optimal tariffs on imports from the same region than on imports from the other region; $t_{ir}^{*}\left(b_{ir},b_{inr};d\right)>t_{inr}^{*}\left(b_{ir},b_{inr};d\right)$.[17]

The reason for adopting the derivation of optimal tariffs in Proposition 1 is for theoretical consistency with the past literature and in particular with our 'benchmark approach' taken by Yi (1996). However, the analytical approach taken in the present paper may be

[16] In Yi's model, under the specification of homogeneous products, his preference function replicates the expression for u_i in the present paper, (2.2). In the model of the present paper, if we let $d=0$ and $k=b_{ir}+b_{inr}$ and $e-c=1$ then $t_{ir}^{*}=t_{inr}^{*}=\left(1+2k\right)/\left(8+3k+2k^{2}\right)$. If we set $n=6$ in Yi's expression for the optimal tariff, presented in *his* Proposition 1, we obtain $\tau\left(k\right)=\left(1+2k\right)/\left(8+3k+2k^{2}\right)$, where $\tau\left(k\right)$ is Yi's notation for the optimal tariff and n is Yi's notation for the number of countries.

[17] For example, consider a two-country TA with one country from each region; say these are countries 1 and 4. Then country 1 sets tariffs $t_{12}=t_{13}=t_{ir}^{*}\left(1,1;d\right)>t_{15}=t_{16}=t_{inr}^{*}\left(1,1;d\right)$ and country 4 sets tariffs $t_{45}=t_{46}=t_{ir}^{*}\left(1,1;d\right)>t_{42}=t_{43}=t_{inr}^{*}\left(1,1;d\right)$.

seen as unsatisfactory in practical terms because it does not require members of an agreement to set a common external tariff. The Most Favored Nation (MFN) principle of the General Agreement on Tariffs and Trade (GATT), adopted in the Charter of the World Trade Organization (WTO), requires that all members of the WTO set the same tariff on each others' imports. Article XXIV, which sets out the WTO rules on TA formation, defines an exception to the MFN principle in that members may allow entry of imports from other members at preferential rates. (Members must endeavor to remove tariffs completely on imports from other members). Nevertheless, Article XXIV requires that the MFN principle for non-members of the TA be upheld.[18] Under our approach, the TA that we analyze violates the MFN principle in that $t_{ir}^{*}(b_{ir}, b_{inr}; d) > t_{inr}^{*}(b_{ir}, b_{inr}; d)$ when $d > 0$.

It would be straight-forward to address this issue by adding an *MFN tariff constraint*, $t_{ir} = t_{inr}$, to the tariff problem set out in (2.13). This constraint would imply that any TA must be a CU. The resulting MFN tariff would be a weighted average of $t_{ir}^{*}(b_{ir}, b_{inr}; d)$ and $t_{inr}^{*}(b_{ir}, b_{inr}; d)$, where the weight would depend on the values of b_{ir} and b_{inr}. The results that follow (Propositions 2, 3, etc.) do not change qualitatively under the MFN tariff constraint since what drives them is the fact that larger rents may be shifted within a region by a *given* tariff, not that the tariff on imports within a region is higher. However, since under the MFN constraint the common external tariff is a weighted average of the tariffs presented in Proposition 1, it is significantly more cumbersome to work with. So we will base the analysis on the discriminatory tariffs presented in Proposition 1.

Article XXIV also stipulates that members may not raise tariffs on non-members when they form or expand a TA. It is straight-forward to check that $t_{ir}^{*}(b_{ir}, b_{inr}; d)$ and $t_{inr}^{*}(b_{ir}, b_{inr}; d)$ are decreasing in agreement size, so Article XXIV is satisfied in this respect.[19]

[18] CU formation is consistent with MFN since, by definition of a CU, external tariffs must be common.

[19] The results of this paper also hold under the more analytically straightforward but less interesting assumption that external tariffs are set at exogenously specified 'MFN' (i.e. common non-prohibitive) levels. The underlying assumption would be that tariffs were pre-determined by multilateral tariff reductions and that TA formation were taking place in that context. The WTO's Article XXIV would then impose a binding constraint that precluded countries from raising average external tariff rates. Goto and Hamada (1999), Syropoulus (1999) and Mrazova, Vines and Zissimos (2009) analyze TA formation under Article XXIV. In summary, the results of the present paper concerning coordination on a regional agreement do not depend on the assumptions made regarding how external tariffs are set providing that the regime does not induce members to compete with each other for third markets. This latter possibility will be discussed in due course.

2.4. Demand functions by region and TA membership

We can now use equilibrium tariffs $t^*_{ir}(b_{ir}, b_{inr}; d)$ and $t^*_{inr}(b_{ir}, b_{inr}; d)$ in (2.7)-(2.10) to write down expressions for equilibrium outputs produced for country i:

$$x_{irm}(b_{ir}, b_{inr}; d) = \frac{2(1 + b_{ir} + b_{inr})(e - c) + (3(1 + b_{ir}) + 2b_{inr}((b_{ir} + b_{inr} - 1)))d}{\Delta(b_{ir}, b_{inr})}; \quad (2.14)$$

$$x_{inrm}(b_{ir}, b_{inr}; d) = \frac{2(1 + b_{ir} + b_{inr})(e - c) - (5 + 2b^2_{ir} + b_{inr}(5 + 2b_{ir}))d}{\Delta(b_{ir}, b_{inr})}; \quad (2.15)$$

$$x_{irnm}(b_{ir}, b_{inr}; d) = \frac{2(e - c) + (3 + b_{inr}(3 + 2(b_{ir} + b_{inr})))d}{2\Delta(b_{ir}, b_{inr})}; \quad (2.16)$$

$$x_{inrnm}(b_{ir}, b_{inr}; d) = \frac{2(e - c) - (5 + b_{ir}(3 + 2(b_{ir} + b_{inr})))d}{2\Delta(b_{ir}, b_{inr})}. \quad (2.17)$$

It can be seen by inspection that trade flows are lowest between countries that are not members of the same TA and are not in the same region; $x_{inrnm}(b_{ir}, b_{inr})$ is the smallest of the quantities given by (2.14)-(2.17).[20] Also, by (2.17), $x_{inrnm}(b_{ir}, b_{inr})$ is decreasing in d. It follows that, by placing an upper bound on d, we can ensure that $x_{inrnm}(b_{ir}, b_{inr}) > 0$ and that in turn all trade flows are positive. The next result identifies the upper bound on d.[21]

Lemma 2. *Fix $e > c$. If $d \in (0, (e - c)/22)$ then, for $b_{ir} \in \{1, 2, 3\}$, $b_{inr} \in \{0, 1, 2, 3\}$ and $b_{ir} + b_{inr} \leq 5$ we have that $x_{irm}(b_{ir}, b_{inr}) > x_{inrm}(b_{ir}, b_{inr}) > x_{irnm}(b_{ir}, b_{inr}) > x_{inrnm}(b_{ir}, b_{inr}) > 0$ and $t^*_{ir}(b_{ir}, b_{inr}) > t^*_{inr}(b_{ir}, b_{inr}) > 0$.*

To restrict attention to positive output levels and positive optimal tariffs, the following standing assumption will be imposed throughout.

Assumption 1. $d \in [0, (e - c)/22)$.

Thus, TA formation always entails the removal of positive tariffs.

[20] Henceforth, the parameter d will be dropped from functional notation so that, for example, $t_{ir}(b_{ir}, b_{inr}; d)$ will be written $t_{ir}(b_{ir}, b_{inr})$ and $x_{inrnm}(b_{ir}, b_{inr}; d)$ will be written $x_{inrnm}(b_{ir}, b_{inr})$.

[21] The reason for restricting attention to $b_{ir} + b_{inr} \leq 5$ in Lemma 2 is because there are no non-regional non-members under free trade ($b_{ir} = 3$, $b_{inr} = 3$), and so it does not make sense to calculate a quantity for $x_{inrnm}(3, 3)$. Using $b_{ir} + b_{inr} \leq 6$ instead would not affect the results qualitatively.

2.5. TA Expansion and Welfare

In this subsection, we will look at the effect of (exogenously specified) TA formation and expansion on member and non-member welfare. We will follow Yi (1996) by looking first at the effect of TA formation on non-member countries. Yi shows (in his Proposition 3) that if a TA forms or expands, then non-member countries are adversely affected. We will now show that Yi's result extends directly to the present model.

TA expansion may occur within a region (in which case b_{ir} increases) or across regions (in which case b_{inr} increases). Thus, define *TA expansion* as an increase in b_{ir} and/or b_{inr}.[22] *TA formation* is just a special case of TA expansion in which all members of the TA that forms start as singletons.

Also note that TA expansion only affects non-members through the demand for exports. This is because optimal tariff setting of non-members is unaffected by TA formation. Thus we can evaluate the effect of TA formation on non-members entirely in terms of the effect on non-member exports to the TA, x_{irnm} and x_{inrnm}, and hence export profits.

Proposition 2. *A non-member country's volume of exports and export profits to a TA of size b_{ir}, b_{inr} is decreasing in b_{ir} and decreasing in b_{inr}. The expansion or formation of a TA reduces the welfare of non-member countries.*

As a TA expands, and removes internal trade barriers, demand for X by consumers in member countries turns towards TA members and away from non-members, hurting the export profits of non-members. This result accords with Bond and Syropoulos (1996) in a perfectly competitive framework and Yi (1996) in an oligopolistic framework.

Others have obtained the opposite result, that TA expansion benefits outsiders; two key examples are Bond, Riezman and Syropoulos (2004, an endowments model) and Ornelas (2005a, an oligopoly model). The key to this discrepancy lies in the fact that in the present paper, as in Yi (1996), countries maximize welfare jointly when setting tariffs while in Bond, Riezman and Syropoulos (2004) and in Ornelas (2005a) countries maximize individual welfare while they agree to remove mutual tariffs. In other words, Bond and Syropoulos (1996) and

[22] Say that a TA initially has two members, one in each region. Then say that one member breaks up with its partner and instead forms a TA with two countries from its own region. Although a new larger TA is created, this is not allowable under our definition of TA expansion since it involves cessation/contraction of membership of the initial TA.

Yi (1996) study CUs while the latter two papers study FTAs. (Recall from Section 2.3.1 that under the MFN constraint, the present model would generate CUs as well.)

In an oligopoly model the discrepancy in outcomes can be attributed directly to a difference in the way export profits are treated by governments. (Essentially the same effect operates in an endowments model through the effect of tariff changes on the terms-of-trade, but cannot be attributed to firm profits since the model does not admit this channel.) A CU common external tariff maximizes the collective profit of all CU member exports to the rest of the world. The same feature is present in our model. In an FTA, by contrast, countries do not care about each others' export profits and hence compete in tariffs for third markets. Thus, while tariffs in the present model fall with TA expansion, they would fall more under FTA expansion, so much so that trade flows would increase not just between agreement members but between non-members as well. Thus both types of TA formation tend to hinder the move to free trade: for CU formation this is so because members gain at non-members' expense and therefore do not wish to see them join; for FTA formation this is because non-members free-ride on the agreement, doing better by remaining outside it than by joining it. Our analysis will be based on the former set of interactions.[23]

Let us now examine the effect of TA formation on the welfare of members. Yi shows for his model that the *joint* welfare of countries involved in TA expansion increases (where 'joint' implies the welfare of existing members and new members). And more generally, if several TAs merge to form a larger TA the aggregate welfare of the member countries increases. Yi remarks that consumer surplus displays a non-monotonicity that is present in underlying optimal external tariffs; the consumer surplus in member countries may first decrease and then increase as a TA expands. A country's export profits, on the other hand, may initially increase but ultimately decrease as the TA expands. The present model introduces a further ambiguity because there are two common external tariffs; the one levied on countries in the same region and the one levied on countries in the other region. Even though the economic environment is made more complicated by the regional dimension of the model, the next result shows that Yi's Proposition 4 extends to the present setting as well.

[23]The latter set of interactions are beyond the scope of the present paper but it would be interesting to study these in a regional setting in future research.

Proposition 3. *The expansion or formation of a TA increases the aggregate welfare of member countries.*

If a set of countries abolishes tariffs internally and sets external tariffs to maximize aggregate welfare then their joint welfare must improve. Proposition 3 shows that the formation of a TA improves joint welfare of member countries even if non-negative tariffs on imports are the only policy tools and even though members and non-members may be in different regions.

So far, we have seen that Yi's results concerning TA expansion in an environment where all countries are identical extend to the present setting where countries may differ by regional location. When a TA expands, this increases the aggregate welfare of the countries in the TA and harms countries that are not members of the TA. Therefore, just as in the world where countries are identical, this implies that the effect of TA expansion on global welfare is ambiguous. The single case in which this ambiguity disappears is the case where TA expansion goes all the way to the grand coalition, which is equivalent to world free trade. Thus, Yi's Proposition 5 carries over to the present setting and is reproduced here for completeness.

Proposition 4. *The effects on global welfare of the formation or expansion of TAs are ambiguous, except when the grand TA forms. World welfare is higher under the grand TA (world free trade) than under any other TA structure.*

All of Yi's results that we have examined so far extend to the present setting. These results have focused on the welfare effects of TA expansion on non-members and on the aggregate welfare of members.

Let us now focus explicitly on the welfare of individual member countries in the TA formation process. In doing so, we will show that a key property of Yi's identical-country model *fails* to hold when transport costs are sufficiently large but still in the range where trade flows between all countries are positive. Of course, Yi's result continues to hold when transport costs are sufficiently small.

Proposition 5. *There exists a unique value $d' \in (0, (e-c)/22)$ such that for $d \in [0, d')$, a country is better off in a (4-country) TA consisting of itself and all 3 countries* in the other region *than in a (3-country) regional TA* in its own region. *For $d \in [d', (e-c)/22)$, a country is better off in a (3-country) regional TA in its own region than in a (4-country) TA consisting of itself and all 3 countries in the other region.*

For $d \in [0, d')$ this result is consistent with Yi's Proposition 8, which says that a member of a TA becomes better off if it leaves its TA to join another TA of equal or larger size. But for $d \in [d', (e-c)/22)$, our result says that a country is better off remaining in a 3-country TA within its own region than it would be if it left its regional TA to form a 4-country TA with all three countries in the other region.[24]

To understand the intuition behind this result, let us consider a member of a regional TA (in its own region), and ask whether it could gain by joining a regional TA in the other region. Say that country 1 is initially in a regional TA; $1 \in B_1 = R_1$. And say that the countries in the other region form another regional TA, $B_2 = R_2$. Country 1 considers whether it could gain by leaving B_1 to join B_2. Decompose the process into three steps: (i) Original members of B_2 abolish tariffs on imports from country 1 and change tariffs on the other countries in R_1 from $t^*_{inr}(3,0)$ to $t^*_{inr}(3,1)$; (ii) Country 1 abolishes tariffs on all countries in B_2, and levies tariffs at $t^*_{ir}(1,3)$ on its two former TA partners in B_1; (iii) The remaining two members of B_1 change tariffs on the (original) members of B_2 (who are located in R_2) from $t^*_{inr}(3,0)$ to $t^*_{inr}(2,0)$ and levy a tariff $t^*_{ir}(2,0)$ on country 1.

Consider the effect of each of these steps on the welfare of country 1 for $d \in [0, d')$ and $d \in [d', (e-c)/22)$ respectively. Take $d \in [0, d')$ first. (i) The abolition of tariffs by the members of B_2 has a positive impact on the welfare of country 1, because country 1 enjoys greater openness in three markets. (ii) Country 1's abolition of tariffs on all three countries in B_2 also improves welfare but the implementation of tariffs on its two former TA partners in B_1 reduces welfare; the net effect is *positive* because access is increased in three markets while it is reduced in only two. (iii) Finally, the implementation of tariffs by its two former TA partners in B_1 reduces export profits and hence welfare in country 1. But the effect on exports of increased access to its three new partners in B_2 (in step (i)) more than

[24]We are assuming that if a country is just indifferent between forming a regional agreement or a non-regional agreement then it exhibits a preference for the regional agreement. This assumption is trivial, and could be reversed without consequence.

compensates. The positive effect on consumer surplus from net tariff removal in moving to the larger TA is greater than the negative effect on tariff revenue and the loss of domestic profits from greater competition in the domestic market.

Now take $d \in [d', (e-c)/22)$. The impact on welfare for country 1 of moving from B_1 to B_2 is reversed. (i) As before, the removal of tariffs by country 1's three new partners in B_2 has a positive impact on export profits. (ii) Once again, country 1's abolition of tariffs on all three countries in B_2 improves welfare while the implementation of tariffs on its two former TA partners in B_1 reduces welfare. But in the presence of transport costs, the net effect is *negative* because the implementation of tariffs by two nearby partners has a larger negative effect on export profits than the removal of tariffs by the three new distant partners in the other region. (iii) Again, the implementation of tariffs by its two former TA partners in B_1 reduces export profits and hence welfare in country 1. And now, the effect on exports of access to its three new partners in B_2 (in step (i)) is not sufficient to compensate. The positive effect on consumer surplus from net tariff removal in moving to the larger TA is smaller than the negative effect on tariff revenue and the loss of domestic profits from greater competition in the domestic market.

Thus, a key result of Yi's is overturned in the present model with the introduction of transport costs. This is significant because it shows that a country will not leave a TA in its own region to form or join a TA in the other region, even if the new TA that forms is larger. In Yi's characterization of an equilibrium TA structure, based on the property that a country would always leave a TA to join a larger one, the first TA to form is always the largest and the last is always the smallest. Proposition 5 therefore calls into question whether, in a regional setting, the TAs that form in equilibrium are necessarily asymmetric in size.

One is bound to ask whether the tendency towards regionalism presented in this result is specific to the model we are using here. Interestingly, Egger and Larch (2008) show that exactly the same effect prevails in a generalization of Krugman's (1991) constant-elasticity-of-substitution model of regionalism. Egger and Larch (2008) have three regions, each of which has two countries. They present simulations in Figure 4 of their paper to show that, with relatively high intercontinental transport costs, a country would rather form a (two country) regional TA than form a (three country) TA by joining a TA with two countries from another region. Bergstrand, Egger and Larch (2009) demonstrate regional TA formation in a different setting where countries are located on a circle and trade costs between countries are increasing with distance. There is no rent shifting in their model, coordination failure is ruled out by assumption, and competition is monopolistic rather than oligopolistic. This suggests that the tendencies towards regionalism derived in the present model extend to other settings as well.[25]

A natural question to ask next is whether the members of a regional TA would invite a country from the other region to join them. The next result shows that, once again, the answer depends on the size of transport costs.

Proposition 6. *There exists a value $d'' \in (d', (e-c)/22)$ such that for $d \in [0, d'')$ the highest feasible level of welfare is achieved when a country is a member of a TA with all of its regional partners plus one country from the other region while non-members are singletons. For $d \in [d'', (e-c)/22)$, the highest feasible level of welfare is achieved when a country is a member of a regional TA (with all members from its own region and no members from the other region) while non-members are singletons.*

This result is again in keeping with Yi (1996). A group of countries can obtain a higher level of welfare than under free trade by forming a TA while non-members remain as singletons. In Yi's model (with six countries and homogenous goods) the highest level of welfare is achieved by a country when it forms a TA of four members. This continues to be true in our model for $d \in [0, d'')$, i.e. when transport costs are small. When transport costs are larger, i.e. for $d \in [d'', (e-c)/22)$, a country does better by forming a regional TA (only with members from its own region). The reason is that the terms-of-trade benefits of TA formation increase with transport costs, and these benefits are increasing in the number of countries left outside the TA. In either case, to maximize national welfare, the TA of which a country is a member must include all of its regional partners.

We are now in a position to see why it would not have been an option for us to adopt Yi's approach of using Bloch's (1996) 'size announcement' game to model TA formation in the present regional setting.[26] The application of that game to the present framework would

[25] Egger and Larch (2008) identify these effects to make sense of their empirical investigation of tendencies towards regional TAs. Bergstrand, et al (2009) examine, both theoretically and empirically, the timing of TA formation and enlargement.

[26] See Bloch (1996) for further discussion of equilibrium existence issues when players are not identical.

be the following. All countries are placed on a list, say 1, 2, ..., 6. Country 1 would be asked to announce the size of the agreement that it would like to form. Then, all proposed partners (following subsequently from country 1) would be asked to agree or disagree. If a proposed partner disagrees then it is asked to make its own proposal of a TA and, again, each subsequent proposed partner is asked whether or not it agrees. If all agree then those countries withdraw from the game, and the next country on the list is asked to announce the size of the TA that it wants to form. If the end of the list of countries is reached then there is a return to the first country on the list that has not already formed an agreement and withdrawn from the game. A stationary equilibrium is reached when all countries have withdrawn from the game.

Now consider what would happen if the size announcement game were played based on our model for $d \in [0, d'')$. By Proposition 6, country 1 would announce that it wants to form a 4-country TA consisting of itself and countries 2, 3 and 4. But, by Proposition 5, country 4 would do better in a TA with all of its regional partners so it refuses (while countries 2 and 3 accept). When country 4 is asked to make an alternative proposal, by Proposition 6, it proposes a TA consisting of itself and countries 5, 6 and 1. This is a mirror of country 1's original proposal. It is now clear that no equilibrium would exist in this situation.[27] In addition to providing a way to capture the coordination problem in TA formation, the TA formation game presented in the next section also provides a way around this equilibrium existence issue.

3. The TA Formation Game

As argued in the Introduction, a country has many potential options for partners when seeking a TA, and this creates potential for coordination failure. We will capture this problem formally by basing the TA formation process on the δ-game of Hart and Kurz (1983). We will now set out the salient features of the TA formation game and how these reflect aspects of the TA formation process that we want to focus on.

Hart and Kurz (1983) consider four simple coalition formation games, denoted by α, β, γ and δ. The common element in all of these games is the selection procedure by which each

[27]For $d \in [d'', (e - c)/22)$, an equilibrium does exist for the size announcement game in which two regional TAs form, one in each region.

and every player simultaneously and without communicating writes down a list of the other players with whom she would like to form a coalition. After the players have completed the selection procedure, they are able to observe the membership of the coalitions that will potentially form. The difference between the four games is the notion of stability; the description of what happens if one or more player wishes to leave the proposed coalition having observed the potential membership. For our purposes, an appealing feature of the δ-game is that if, having learned the identity of coalition members, any player wishes to leave a coalition she can do so while the coalition otherwise remains in tact.[28] This captures a feature of TA formation, which seems to reflect actual practice, that if one proposed partner chooses not to go ahead with the TA then the others may still do so.[29]

A second feature of actual coalition formation that we will capture here is that countries who approach each other to form a TA have more information about the prospective membership of their agreement than countries who have not approached each other. In the TA formation game, we will adopt a particularly tractable form of this assumption; until the TA formation game is complete, each country only finds out about the prospective TA membership of its own TA partners. If a country does not observe others to have joined its own TA then it assumes they have not formed one. We adopt this approach from Arnold and Wooders (2005), who introduce this informational friction in their formalization of club formation.[30]

This assumption replaces the assumption made by Hart and Kurz (1983) that each country has the same information about the coalition formation activities of all others. In an abstract setting the approach of Hart and Kurz is undoubtedly more appealing. Yet in the present applied setting our assumption seems to capture an important aspect of the informational frictions that are likely to affect the actual process of TA formation.[31]

[28]In the γ-game, by contrasting example, all players in a proposed coalition return to singleton status if one player leaves. The α-game and β-game embody similar variations in coalition stability.

[29]For example, even though Britain dropped out of the original discussions to form the European Economic Community (which later became the EU) the original signatories to the Treaty of Rome still went ahead in 1957.

[30]Arnold and Wooders (2005) use the word 'club' in the same sense as Hart and Kurz (1983) use the word 'coalition'. Both concepts correspond to the notion of TA as we use it here.

[31]Even if a country can deduce that it is in the national interests of another group of countries to form a TA in response to its own TA formation activities, the assumption will be that it does not act on the deduction. An interpretation of this assumption is that countries attach a degree of 'political uncertainty' to the TA-formation process due to possible distributional or politically-motivated concerns of policy-makers which cannot be observed from outside the country.

In formal terms, the informational friction serves to introduce a degree of stability and consequently enlarges the set of possible equilibria. After we have shown that the set of equilibria is potentially large, we will then show how the introduction of a second friction, that of transport costs, reduces the number of equilibria to one.

The process of TA formation is initialized with a TA structure in which there are *no TAs*; initially the TA structure, B, is the set of singletons. At that point, each country i chooses a *strategy*, s_i, where each s_i contains a list of countries in N with which country i would like to form a TA; this list includes country i itself.[32] The *strategy space*, S_i, for country i is the set of all subsets of N, i.e. the set of all possible TAs that could include country i. Strategies are chosen simultaneously. After strategies are chosen, a country's prospective TA partners are revealed to it. A country does not observe the strategies of other countries. If a country does not observe another country as its TA partner, it maintains the assumption that the other country is a singleton.

A *bilateral trade accord* (i, j) is formed if and only if $i \in s_j$ and $j \in s_i$. A subset of countries B_k is a *TA* if and only if all pairs of countries in B_k have a bilateral trade accord. This assumption ensures that a TA forms if and only if there is unanimous support for its membership. If a country finds itself in the position of being in two or more otherwise exclusive and otherwise unanimous TAs, it chooses the TA that maximizes its payoff under the assumption that the memberships of the TA that it joins and the TA that it leaves remain otherwise constant.[33] When a country chooses one TA over another one, it assumes that the other goes ahead without it.

Each strategy vector $s = (s_1, ..., s_N)$ *induces* a unique TA structure, B, and so we can now write B as a function of s; $B(s)$:

$$B(s) = \{ (i, j) | \, i \in s_j, j \in s_i \} .$$

Since a TA structure implies a unique value of b_{ir} and b_{inr} for each country i, and since these

[32] The purpose of including i in s_i is that then we can view B_k as the intersecting set of all the elements of strategies s_i for all $i \in N$.

[33] Pushing this one step further, any two countries caught between two TAs will assume that each behaves in the same way as the other in the TA that they choose. This assumption is the same as that of Hart and Kurz, that if any player is caught between two coalitions then it chooses the biggest one under the assumption that all other players caught in the same situation do the same. In a world where all countries are identical this assumption is innocuous. In principle this assumption could lead to mistakes in a world where countries differ but this potential problem will not be an issue for any of the situations that we will study.

in turn imply values of $t^*_{ir}(b_{ir}, b_{inr})$ and $t^*_{inr}(b_{ir}, b_{inr})$, the payoff to country i associated with s can be represented simply as $w_i = w_i(t_{ir}(B(s)), t_{inr}(B(s)))$; the payoff for country i from the TA structure *induced by* s. For compactness, we may write $w_i = w_i(s)$.

The notion of equilibrium is adapted from Arnold and Wooders (2005). For any given TA structure $B = (B_1, ..., B_k, ..., B_m)$, a strategy vector $s^* \in S$ is a *Nash club equilibrium* of the TA formation game if for any given $B_k \in B$ there is no $Z \subseteq B_k$ and $s \in S$ such that

1. $s_i = s^*_i$ for all $i \notin Z$.
2. $w_i(s) \geq w_i(s^*)$ for all $i \in Z$ and $w_i(s) > w_i(s^*)$ for some $i \in Z$.

By definition, an equilibrium exists if no group of countries Z in some TA, B_k, can do better by deviating, i.e. by forming a TA of their own. In formal terms the difference between our assumption and that of Hart and Kurz may be understood as follows. Hart and Kurz allow deviations to be undertaken by *any* coalition $Z \subseteq N$ (in contrast to our restriction of deviations to $Z \subseteq B_k$). Thus, our definition weakens the notion of equilibrium relative to Hart and Kurz, admitting a relatively large number of equilibria. In particular, it does not exclude from the equilibrium set candidates that arise as a result of coordination failure - in the present context, where countries could all benefit by merging their proposed TAs but fail to do so due to the informational friction. It remains to be shown how the problem of coordination arises when all countries are identical. We will then show that it is resolved, albeit inefficiently, when countries may differ by region.

4. The Problem of Coordination Failure

We will now show how the problem of coordination arises in a world where all countries are identical. To do so, we will fix $d = 0$. By Proposition 6, we know that a TA of four countries maximizes the welfare of its members (if the other two countries are singletons). The problem of coordination failure arises because, even if each country writes down a strategy s_i with four elements, in the absence of communication there are many possible TA structures that may arise in equilibrium as a result of all countries playing this strategy. An equilibrium may arise in which there is a TA with four countries, which is the desired outcome of each of the members. But of course the two countries excluded from the four-country TA do not achieve their desired outcome. Moreover, this is not the only TA structure that can be sustained in equilibrium. We will first consider an equilibrium in which there is a four-country TA, but then consider one of many possible alternative TA structures that may arise when all countries seek to form a TA with four members.

4.1. Various equilibria with coordination failure

An example of a strategy vector, s, that gives rise to an equilibrium in which there is a four-country TA is as follows:

$$s_1 = s_2 = s_3 = s_4 = \{1, 2, 3, 4\}\,,\ s_5 = s_6 = \{1, 2, 5, 6\}\,.$$

Notice that the strategies $s_1 ... s_4$ form an intersecting set of elements $\{1, 2, 3, 4\}$ while 5 is only listed in s_6 (and s_5 of course) and 6 is only listed in s_5 (and s_6). Thus, the resulting trade agreement structure is $\{\{1, 2, 3, 4\}, \{5, 6\}\}$. It is easy to check that no country can gain by deviating from this agreement structure and so therefore this must be an equilibrium. Consider the allowable deviations. If a member of the four-country TA were to veto membership of another single member then the TA structure would become one of a three-country TA, a singleton and a two-country TA, for example $\{\{1, 2, 3\}, \{4\}, \{5, 6\}\}$. Then, since welfare is maximized in a TA of four countries, the payoff to the country that undertook the veto would fall, as would the payoff of the ejected member. The welfare of 5 and 6 actually increases. If more than one country's membership is vetoed, it is easy to check that the payoff of remaining members falls even further. Therefore, no member of the four-country TA has an incentive to deviate. The same is true for the two-country TA. Thus we have a Nash club equilibrium.

We have already discussed above the reasons why TA member welfare changes when one or more countries are ejected. Let us briefly review why non-member welfare changes. We just noted that, from an initial trade agreement structure of $\{\{1, 2, 3, 4\}, \{5, 6\}\}$, if country 4 is ejected, leaving a trade agreement structure of $\{\{1, 2, 3\}, \{4\}, \{5, 6\}\}$, then the welfare of 5 and 6 increases. Why does this happen? Tariffs set by 5 and 6 do not change because these depend only on their own TA structure, which has not changed. When 4 is ejected, countries 1, 2 and 3 restore tariffs against it, and as a result demand less of X from 4, shifting some of their demand towards 5 and 6. With all else equal, this puts the trade accounts of countries 5 and 6 into surplus, requiring an improvement in their terms-of-trade to restore equilibrium. This adjustment occurs within the model via an increase in the flow of profits to the firms in 5 and 6. In addition, 4 restores tariffs against countries 1, 2 and 3, shifting its demand for X towards 5 and 6. Both of these effects combine to shift profits towards 5 and 6, thus increasing welfare.

Notice that, because $d = 0$, the partition of countries into regions has no relevance to this equilibrium. As specified, the equilibrium contains three countries from R_1 and one country from R_2. But under an equivalent characterization of equilibrium we could have permuted the countries in such a way that two countries were in R_1 (say 1 and 2), and two countries were in R_2 (say 3 and 4). This is due to the fact that all countries are identical. We shall see that the partition of countries into regions does become relevant for equilibrium when $d > 0$.

Now let us consider another possible equilibrium in which there are three TAs, each with two members. This equilibrium arises if each country proposes to form a four-member-TA consisting of itself and the three countries 'next to it':

$$\begin{aligned} s_1 &= \{1,2,3,4\}\,,\ s_2 = \{2,3,4,5\}\,,\ s_3 = \{3,4,5,6\}\,, \\ s_4 &= \{4,5,6,1\}\,,\ s_5 = \{5,6,1,2\}\,,\ s_6 = \{6,1,2,3\}\,. \end{aligned}$$

By inspection of the strategy vector, the agreements that form are $\{1,4\}$, $\{2,5\}$ and $\{3,6\}$. Again, it is straight-forward to check that this is an equilibrium strategy vector. If any member of a two-country agreement vetoes membership of the other, splitting the agreement into two singletons, then its payoff falls by (the reverse of) Proposition 3. This is the only feasible deviation.

5. Transport Costs and Coordination

The problem of coordination failure identified in the previous section is resolved in the presence of a transport cost $d \in (0, (e-c)/22)$. Note that the transport cost may be arbitrarily small.

Proposition 7. *Assume $d \in (0, (e-c)/22)$. There is a unique equilibrium with two regional TAs; $B_1 = R_1$, $B_2 = R_2$. The payoff to each country is the same and is lower than free trade.*

There are two cases to consider, although the outcome is the same in both; one where $d \in (0, d'')$ and one where $d \in [d'', (e-c)/22)$. The second case is easier so we consider that first. By Proposition 6, due to higher transport costs $d \in [d'', (e-c)/22)$, each country anticipates obtaining the highest level of welfare from a regional TA with only the two other countries in its own region. Thus, it is immediate that the intersecting sets formed by countries' strategies is two regional TAs; $B_1 = R_1$ and $B_2 = R_2$.

The case where $d \in (0, d'')$ is slightly more subtle. In that case, each country's welfare is maximized by a 4-member TA with three members from its own region and one member from the other region. But even if all countries write down a strategy containing four countries, three from its own region and one from the other region, the intersecting sets of countries formed by these strategies give rise to two regional TAs; $B_1 = R_1$ and $B_2 = R_2$. To see why, consider the following strategy vector:

$$s_1 = s_2 = s_3 = \{1, 2, 3, 4\}\,, \; s_4 = s_5 = s_6 = \{1, 4, 5, 6\}\,.$$

The strategies $s_1 ... s_3$ form an intersecting set of elements $\{1, 2, 3\}$ and the strategies $s_4 ... s_6$ form an intersecting set of elements $\{4, 5, 6\}$. Thus, $\{\{1, 2, 3\}, \{4, 5, 6\}\}$ is the resulting TA structure. Even though, for this example, all countries that end up in B_1 list country 4, country 4 only names country 1 and *not* 2 and 3. Only the membership of 1, 2 and 3 is unanimous among all members. It is straight forward to check that the same is true for all other possible strategy vectors.

No country can gain by deviating from this agreement structure and so therefore this must be an equilibrium. Consider the allowable deviations. If a member of one of the regional agreements were to veto membership of another single member then the agreement structure would become one of a two-country agreement, a singleton and a three-country agreement; for example $\{\{1, 2, 3\}, \{4, 5\}, \{6\}\}$. Then the payoff to the country that undertook the veto, in this example country 4 or 5, would fall. The welfare of countries in the regional trade agreement that remains, $\{1, 2, 3\}$, increases. As before, if more than one country's membership is vetoed, it is easy to check that the payoff of the remaining member falls even further. Thus, no member of a regional agreement has an incentive to deviate. No deviation is available to the singleton. Thus we have a Nash club equilibrium. This is the only possible equilibrium that can arise for transport costs in the interval $d \in (0, (e-c)/22))$. In equilibrium the TA structure is symmetrical, so each country receives the same payoff. By

Proposition 4, the payoff that each country receives must be lower than under free trade.

Clearly, this outcome depends on the informational friction. If countries had complete information about each other and were far-sighted then each would anticipate that the countries of the other region would form a TA as well. Then each country would be able to see that a move to free trade would be more beneficial. But we can also see how the present assumption regarding informational frictions captures aspects of uncertainty that are likely to be present in the actual process of TA formation across regions.

6. Conclusions

The purpose of this paper has been to show that problems of coordination failure in the formation of TAs may be resolved when countries are organized into regions. Costs of shipping goods between regions are only required to be positive and can be infinitesimally small. With no transport costs, there is a problem of multiple equilibria due to coordination failure familiar from the theory of coalition formation. Positive transport costs are enough to bring about a unique equilibrium in the first period of the TA formation game. Starting from a situation where there are no TAs, two regional TAs form simultaneously.

Inevitably, the theoretical framework developed here simplifies the situation in a number of key respects. The underlying economic structure of the model is one of Cournot competition in a homogeneous product. In practice, the forces of competition are understood to be more subtle and complex. Future research could take steps to see the extent to which the insights of the present model extend to alternative settings. It seems reasonable to argue that the features of our model which drive regional TA formation would extend to other forms of competition. In particular, it is widely appreciated that Bertrand competition behaves like Cournot competition when firms must pre-commit to quantities. And, as suggested by Bond (2001), a more elaborate modeling of perfect competition should exhibit the same features. The key motivating feature of the model is that, in the absence of an agreement, there would be greater rent shifting within a region than across regions. This feature of the model is motivated by the presence of transport costs and will be robust to alternative assumptions about competition between firms and tariff setting between governments. Of course, in the real world, the size of countries and their regional structure is more complex as well. A promising way to address such complexities would be to extend the framework of the present paper to incorporate more sophisticated and realistic simulations-based models of the kind developed by Whalley (1985).

It also seems reasonable to argue that the features of the model would extend to a more elaborate model of production. A direct way to make such an extension would be to assume that X is horizontally differentiated, extending preferences and production accordingly. Alternatively, Syropoulos (2002) offers a way to investigate whether the insights of the regional model developed in the present paper could be extended to a Heckscher-Ohlin framework. This would take explicit account of how differences in factor endowments would interact with countries' organization into geographical regions.

A question that arises through our analysis is how the results would be affected by the trade regime. Our results are based on the assumption that countries coordinate on the setting of external tariffs as in a CU (except that TAs do not necessarily set common external tariffs in our framework). Outcomes might look different if countries were to form FTAs instead. In real life it is CUs that appear most often to be regional, the most prominent examples being the EU, MERCOSUR and the Southern African Customs Union (SACU). All are formed between countries that have contiguous borders. On the other hand, the recent proliferation of FTAs has included countries which are not close to each other, such as the aforementioned FTA between South Korea and the US. This seems to suggest that the forces towards regionalism may operate more forcefully in CUs than in FTAs. This in turn begs the question of what drives the choice of trade regime; the question addressed by Riezman (1999). Thus Riezman presents an approach which could be used to address the question of which trade regime would be adopted within the present regional framework, and if an FTA were adopted whether it would necessarily be regional.

A focus of some research on regionalism is on situations where tariffs are used for political or redistributive purposes.[34] Such considerations could be incorporated in the model of the present paper by putting a heavier weight on producers' profits. It seems possible that producer interests which span regions, as between the UK and the US for example, could counteract the forces towards regionalism identified in our basic framework.

[34] In addition to Krishna (1998) and Ornelas (2005a,b), see for examples Grossman and Helpman (1995) and Maggi and Rodriguez-Clare (1998).

A. Appendix

Proof of Lemma 1. The basic proof strategy was suggested by Monika Mrazova for a different model in Mrazova, Vines and Zissimos (2009). First note that $\pi_{ij} = (x_{ij})^2$ and $\pi_{ji} = (x_{ji})^2$; by (2.5) we have $\pi_{ij} \equiv (p_i - c_{ij})\, x_{ij}$ and by the first order condition of (2.5) we have that $p_i - c_{ij} = x_{ij}$.

Next, the first four terms of (2.11) may be written

$$\begin{aligned} & C_i + T_i + D_i + \pi_{ii} \\ = \; & \frac{1}{2}(e - p_i)\, X_i + \sum_{j \in N} t_{ij} x_{ij} + (x_{ii})^2 \\ = \; & \frac{1}{2} X_i^2 + \sum_{j \in N} t_{ij} x_{ij} + (x_{ii})^2 . \end{aligned}$$

where we have used the fact that $D_i = 0$ by assumption and line 3 uses (2.3). On the other hand, the first four terms of (2.12) may be written

$$\begin{aligned} & v(X_i) - cX_i - \sum_{j \in N} d_{ij} x_{ij} - \sum_{j \in N \backslash \{i\}} (x_{ij})^2 \\ = \; & eX_i - \frac{1}{2} X_i^2 - cX_i - \sum_{j \in N} d_{ij} x_{ij} - \sum_{j \in N \backslash \{i\}} (x_{ij})^2 \\ = \; & eX_i - \frac{1}{2} X_i^2 - cX_i - \sum_{j \in N} d_{ij} x_{ij} - \sum_{j \in N} (x_{ij})^2 + (x_{ii})^2 \\ = \; & eX_i - \frac{1}{2} X_i^2 - cX_i - \sum_{j \in N} d_{ij} x_{ij} - \sum_{j \in N} (p_i - c - t_{ij} - d_{ij})\, x_{ij} + (x_{ii})^2 \\ = \; & eX_i - \frac{1}{2} X_i^2 - cX_i - \sum_{j \in N} d_{ij} x_{ij} - \sum_{j \in N} p_i x_{ij} + cX_i + \sum_{j \in N} t_{ij} x_{ij} + \sum_{j \in N} d_{ij} x_{ij} + (x_{ii})^2 \\ = \; & eX_i - \frac{1}{2} X_i^2 - \sum_{j \in N} p_i x_{ij} + \sum_{j \in N} t_{ij} x_{ij} + (x_{ii})^2 \\ = \; & eX_i - \frac{1}{2} X_i^2 - \sum_{j \in N} (e - X_i)\, x_{ij} + \sum_{j \in N} t_{ij} x_{ij} + (x_{ii})^2 \\ = \; & \frac{1}{2} X_i^2 + \sum_{j \in N} t_{ij} x_{ij} + (x_{ii})^2 . \end{aligned}$$

where the second line uses (2.2), the fourth line uses (2.4) and the first order condition of (2.5), and the seventh line uses (2.3). The result follows. □

Proof of Proposition 1.

Government i 's problem, as expressed in (2.13), simplifies to

$$\begin{aligned}\underset{\{t_{ir},t_{inr}\}_{i\in B_k}}{\text{Max}} \; &(e-c)\,X_i - \frac{X_i^2}{2} - (3-b_{ir})\,(x_{irnm})^2 - (3-b_{inr})\,(x_{inrnm})^2 \\ &\qquad - d\,(b_{inr}x_{inrm} + (3-b_{inr})\,x_{inrnm})\,.\end{aligned}$$

The first order condition with respect to t_{ir} is

$$\begin{aligned}(e-c-1)\,\frac{dX_i}{dt_{ir}} - 2\,(3-b_{ir})\,x_{irnm}\frac{dx_{irnm}}{dt_{ir}} - 2\,(3-b_{inr})\,\frac{dx_{inrnm}}{dt_{ir}} \\ -\left(b_{inr}\frac{dx_{inrm}}{dt_{ir}} + (3-b_{inr})\,\frac{dx_{inrnm}}{dt_{ir}}\right)d = 0.\end{aligned}$$

The first order condition with respect to t_{inr} is

$$\begin{aligned}(e-c-1)\,\frac{dX_i}{dt_{inr}} - 2\,(3-b_{ir})\,x_{irnm}\frac{dx_{irnm}}{dt_{inr}} - 2\,(3-b_{inr})\,\frac{dx_{inrnm}}{dt_{inr}} \\ -\left(b_{inr}\frac{dx_{inrm}}{dt_{inr}} + (3-b_{inr})\,\frac{dx_{inrnm}}{dt_{inr}}\right)d = 0.\end{aligned}$$

Using (2.7)-(2.10) and their first derivatives, a reduced form for each of the above first order conditions may be obtained. Since the objective function is globally concave in t_{ir} and in t_{inr}, there exists a unique symmetric solution for each:

$$\begin{aligned}t_{ir}^* &= \frac{(e-c)\,(1+b_{ir}+b_{inr}) + (24+6b_{ir}-8b_{inr})\,d + (3-b_{inr})\,(15+2\,(b_{ir}+b_{inr}))\,t_{inr}}{2+2b_{ir}^2+b_{ir}\,(9+2b_{inr})+3\,(17-2b_{inr})}; \\ t_{inr}^* &= \frac{(e-c)\,(1+b_{ir}+b_{inr}) - (25-6b_{ir}+8b_{inr})\,d + (3-b_{inr})\,(15+2\,(b_{ir}+b_{inr}))\,t_{ir}}{2+2b_{inr}^2+b_{inr}\,(9+2b_{ir})+3\,(17-2b_{ir})}.\end{aligned}$$

Solving simultaneously for t_{ir}^* and t_{inr}^* obtains the result. $\square$

Proof of Lemma 2. The fact that $x_{irm}\,(b_{ir},b_{inr}) > x_{inrm}\,(b_{ir},b_{inr}) > x_{irnm}\,(b_{ir},b_{inr}) > x_{inrnm}\,(b_{ir},b_{inr})$ is established by inspection of (2.14)-(2.17). It remains to show that if $d \in (0,(e-c)\,/22)$ then $x_{inrnm}\,(b_{ir},b_{inr}) > 0$. Since $x_{inrnm}\,(b_{ir},b_{inr})$, as given by (2.17), is decreasing d, we can solve for the largest value of d at which $x_{inrnm}\,(b_{ir},b_{inr}) = 0$ (for $b_{ir} \in \{1,2,3\}$, $b_{inr} \in \{0,1,2,3\}$ and $b_{ir}+b_{inr} \leq 5$). The solution for the value of d at which $x_{inrnm}\,(b_{ir},b_{inr}) = 0$, denoted by $\widetilde{d}$, is

$$\widetilde{d} = \frac{2\,(e-c)}{5+3b_{ir}+2b_{ir}b_{inr}+2b_{ir}^2}.$$

The solution $\widetilde{d}$ is globally decreasing in b_{ir} and b_{inr}, so use $b_{ir}=3$, $b_{inr}=2$ in the solution to yield $\widetilde{d} = (e-c)\,/22$. It can be checked by substitution that $x_{inrnm}\,(b_{ir},b_{inr}) > 0$ for $b_{ir}=2$, $b_{inr}=3$. The result follows. $\square$

Proof of Proposition 2. It must be established that $dx_{irnm}/db_{ir} < 0$, $dx_{inrnm}/db_{ir} < 0$, $dx_{irnm}/db_{inr} < 0$, and $dx_{inrnm}/db_{inr} < 0$ over the range of feasible b_{ir} and b_{inr}. Each case will be taken in turn. Differentiating $x_{irnm}(b_{ir}, b_{inr})$ with respect to b_{ir}, we obtain

$$\begin{aligned}\frac{dx_{irnm}(b_{ir}, b_{inr})}{db_{ir}} &= \frac{2b_{inr}d}{2\Delta(b_{ir}, b_{inr})} \\ &\quad - \frac{(6+8(b_{ir}+b_{inr}))(2(e-c)+(3+b_{inr}(3+2(b_{ir}+b_{inr})))d)}{(2\Delta(b_{ir}, b_{inr}))^2}\end{aligned}$$

$$= -\frac{2(3+4(b_{ir}+b_{inr}))(e-c)+(9+12b_{ir}+b_{inr}(1+2(b_{ir}+b_{inr}))(5+2(b_{ir}+b_{inr})))d}{2(\Delta(b_{ir}, b_{inr}))^2}.$$

So $dx_{irnm}(b_{ir}, b_{inr})/db_{ir} < 0$ for all $d \geq 0$.

Differentiating $x_{inrnm}(b_{ir}, b_{inr})$ with respect to b_{ir}, we obtain

$$\begin{aligned}\frac{dx_{inrnm}(b_{ir}, b_{inr})}{db_{ir}} &= -\frac{(3+4b_{ir}+2b_{inr})d}{2\Delta(b_{ir}, b_{inr})} \\ &\quad - \frac{(6+8(b_{ir}+b_{inr}))(2(e-c)-(5+b_{ir}(3+2(b_{ir}+b_{inr})))d)}{(2\Delta(b_{ir}, b_{inr}))^2}\end{aligned}$$

$$= -\frac{2(3+4(b_{ir}+b_{inr}))(e-c)+(9+12b_{ir}+b_{inr}(1+2(b_{ir}+b_{inr}))(5+2(b_{ir}+b_{inr})))d}{2(\Delta(b_{ir}, b_{inr}))^2}.$$

So $dx_{inrnm}(b_{ir}, b_{inr})/db_{ir} < 0$ for all $d \geq 0$. (After simplification, we see that $dx_{irnm}(b_{ir}, b_{inr})/db_{ir} = dx_{inrnm}(b_{ir}, b_{inr})/db_{ir}$.)

Differentiating $x_{irnm}(b_{ir}, b_{inr})$ with respect to b_{inr}, we obtain

$$\begin{aligned}\frac{dx_{irnm}(b_{ir}, b_{inr})}{db_{inr}} &= \frac{(3+2b_{ir}+4b_{inr})d}{2\Delta(b_{ir}, b_{inr})} \\ &\quad - \frac{(6+8(b_{ir}+b_{inr}))(2(e-c)+(3+b_{inr}(3+2(b_{ir}+b_{inr})))d)}{(2\Delta(b_{ir}, b_{inr}))^2}\end{aligned}$$

$$= \frac{-2(3+4(b_{ir}+b_{inr}))(e-c)+(15+13b_{ir}+4(5b_{inr}+b_{ir}b_{inr}(3+b_{inr})+(3+2b_{inr})b_{ir}^2+b_{ir}^3))d}{2(\Delta(b_{ir}, b_{inr}))^2}.$$

The second term in the numerator is positive and increasing in b_{ir}, b_{inr} and d while the first term is negative. It is easily checked that overall the numerator is negative for $b_{ir} = 3$, $b_{inr} = 2$ and $d = (e-c)/22$. So $dx_{irnm}(b_{ir}, b_{inr})/db_{inr} < 0$ for all feasible $\{b_{ir}, b_{inr}\}$ pairs and $d \in (0, (e-c)/22)$.

Differentiating $x_{inrnm}(b_{ir}, b_{inr})$ with respect to b_{inr}, we obtain

$$\begin{aligned}\frac{dx_{inrnm}(b_{ir}, b_{inr})}{db_{inr}} &= -\frac{2b_{inr}d}{2\Delta(b_{ir}, b_{inr})} \\ &\quad -\frac{(6+8(b_{ir}+b_{inr}))(2(e-c)-(5+b_{ir}(3+2(b_{ir}+b_{inr})))d)}{(2\Delta(b_{ir}, b_{inr}))^2}\end{aligned}$$

$$= \frac{-2(3+4(b_{ir}+b_{inr}))(e-c)+(15+13b_{ir}+4(5b_{inr}+b_{ir}b_{inr}(3+b_{inr})+(3+2b_{inr})b_{ir}^2+b_{ir}^3))d}{2(\Delta(b_{ir}, b_{inr}))^2}.$$

After simplification, we see that $dx_{inrnm}(b_{ir}, b_{inr})/db_{inr} = dx_{irnm}(b_{ir}, b_{inr})/db_{inr}$ So it must be the case that $dx_{inrnm}(b_{ir}, b_{inr})/db_{inr} < 0$ for all feasible $\{b_{ir}, b_{inr}\}$ pairs and $d \in (0, (e-c)/22)$. □

Proof of Proposition 3. The strategy of proof follows Yi (1996). Assume that there exists a TA structure $B = (B_1, B_2, ..., B_m)$ and that two or more TAs, say B_1, B_2 ... B_r, merge to create an enlarged TA. We will show that the total welfare of the members of the enlarged TA increases. To do this, we will show that the tariff changes required to implement TA enlargement undertaken by any one given member of the enlarged TA must increase the aggregate welfare of all members. Thinking of TA enlargement as a sequence of such tariff changes by each and every member then gives the result.

Claim. Initially, before the merger, country i has free trade with $b_{ir}-1$ countries in its own region and b_{inr} countries in the other region. Country i levies a tariff $t_{ir}(b_{ir}, b_{inr})$ on each of the $3-b_{ir}$ non-members in its own region and a tariff $t_{inr}(b_{ir}, b_{inr})$ on each of the $3-b_{inr}$ countries in the other region. As a result of the merger, in the new enlarged TA, country i shares a TA with $b'_{ir}-1$ countries in its own region and b'_{inr} countries in the other region. Let $h_{ir} = b'_{ir} - b_{ir} \geq 0$ and $h_{inr} = b'_{inr} - b_{inr} \geq 0$. Country i abolishes tariffs on h_{ir} countries in its own region and h_{inr} countries in the other region, and changes tariffs to $t'_{ir}(b'_{ir}, b'_{inr})$ on each of the $3-b'_{ir}$ non-members in its own region and changes tariffs to $t'_{inr}(b'_{ir}, b'_{inr})$ on each of the $3-b'_{inr}$ non-members in the other region. Then the aggregate welfare of the $b_{ir}+h_{ir}+b_{inr}+h_{inr}$ countries in the enlarged TA (which consists of country i, $b_{ir}+b_{inr}-1$ countries who paid no tariffs initially and $h_{ir}+h_{inr}$ countries whose tariffs were abolished) improves.

Proof. Without loss of generality, consider the TA B_1, of which country 1 is assumed to be a member. B_1 has b_{1r} members from R_1 and b_{1nr} members from R_2. Then let membership expand to create an enlarged TA, B_1', consisting of b_{1r}' members in R_1 and b_{1nr}' members in R_2 (where all original members are also members of the enlarged TA). The comparative statics exercise that we will now carry out is as follows. We will calculate the effect on the aggregate welfare of all countries in B_1' that results when country 1 abolishes tariffs on h_{1r} countries in R_1 and h_{1nr} countries in R_2, and changes tariffs on $(3 - b_{1r} - h_{1r})$ non-members in R_1 from $t_{1r}(b_{1r}, b_{1nr})$ to $t_{1r}(b_{1r}', b_{1nr}')$ and on $(3 - b_{1nr} - h_{1nr})$ non-members in R_2 from $t_{1nr}(b_{1r}, b_{1nr})$ to $t_{1nr}(b_{1r}', b_{1nr}')$.

Define

$$\begin{aligned} \Delta t_{1r} &= t_{1r}(b_{1r}, b_{1nr}) - t_{1r}(b_{1r}', b_{1nr}'); \\ \Delta t_{1nr} &= t_{1nr}(b_{1r}, b_{1nr}) - t_{1nr}(b_{1r}', b_{1nr}'). \end{aligned}$$

First consider infinitesimal changes in tariffs

$$d\mathbf{t} \equiv (0, ..., 0, dt, ..., dt, dt_r, ..., dt_r, 0, ..., 0, \phi dt, ..., \phi dt, dt_{nr}, ..., dt_{nr})$$

from a tariff vector

$$\mathbf{t} \equiv (0, ..., 0, t, ..., t, t_r, ..., t_r, 0, ..., 0, \phi t, ..., \phi t, t_{nr}, ..., t_{nr}),$$

where: dt appears from the $(b_{1r} + 1)$th element to the $(b_{1r} + h_{1r})$th element; ϕdt appears from the $(b_{1nr} + 4)$th element to the $(b_{1nr} + h_{1nr} + 3)$th element, unless $b_{1nr} = h_{1nr} = 0$ in which case dt_{nr} appears from the 4th to the last element; dt_r appears from the $(b_{1r} + h_{1r} + 1)$th element to the 3rd element; dt_{nr} appears from the $(b_{1nr} + h_{1nr} + 4)$th element to the last element. The tariff t was already being imposed on new TA members in the same region and is reduced to zero through the TA formation process. The tariff ϕt (i.e. $\phi \times t$) was already being imposed on new TA members from the other region, where $\phi = t_{1nr}/t_{1r}$ (see below for specification of t_{1nr} and t_{1r}). Also,

$$\begin{aligned} dt_r &\equiv \frac{\Delta t_{1r}}{t_{1r}(b_{1r}, b_{1nr})} dt; \\ dt_{nr} &\equiv \frac{\Delta t_{1nr}}{t_{1nr}(b_{1r}, b_{1nr})} \phi dt. \end{aligned}$$

Start from

$$\mathbf{t}(b_{1r}', b_{1nr}') \equiv (0, ..., 0, t_{1r}(b_{1r}', b_{1nr}'), ..., t_{1r}(b_{1r}', b_{1nr}'), 0, ..., 0, t_{1nr}(b_{1r}', b_{1nr}'), ..., t_{1nr}(b_{1r}', b_{1nr}'))$$

where 0 appears from the first to the $(b_{1r}+h_{1r})$th element and from the fourth to the $(b_{1nr}+h_{1nr}+3)$th element (unless $b_{1nr}=h_{1nr}=0$). We can move to

$$\mathbf{t}\left(b_{1r},b_{1nr}\right)\equiv\left(0,...,0,t_{1r}\left(b_{1r},b_{1nr}\right),...,t_{1r}\left(b_{1r},b_{1nr}\right),0,...,0,t_{1nr}\left(b_{1r},b_{1nr}\right),...,t_{1nr}\left(b_{1r},b_{1nr}\right)\right)$$

where 0 appears from the first to the (b_{1r})th element and from the fourth to the $(b_{1nr}+4)$th element (unless $b_{1nr}=0$) by integrating the infinitesimal changes $d\mathbf{t}$ from 0 to $\mathbf{t}\left(b_{1r},b_{1nr}\right)$. Below, we will show that $d\left(\sum_{j\in B_1'}w_j\right)/d\mathbf{t}<0$ for all $\mathbf{t}$ along such a path of integration. The claim then follows.

Since changes in country 1's tariffs do not affect sales in other countries,

$$d\left(\sum_{j\in B_1'}w_j\right)/d\mathbf{t}=d\left(\hat{w}_1+\sum_{j\in B_1'\setminus\{1\}}\pi_{1j}\right)/d\mathbf{t},$$

where $\hat{w}_1$ is country 1's welfare net of its exports. Since

$$\hat{w}_1+\sum_{j\in N\setminus\{1\}}\pi_{1j}=v\left(X_1\right)-cX_1-d\sum_{j\in R_2}x_{1j},$$

it follows that

$$\hat{w}_1+\sum_{j\in B_1'\setminus\{1\}}\pi_{1j}=v\left(X_1\right)-cX_1-\sum_{j\in N\setminus B_1'}\pi_{1j}-d\sum_{j\in R_2}x_{1j}.$$

The proportional relationship between $t_{1r}\left(b_{1r},b_{1nr}\right)$ and $t_{1nr}\left(b_{1r},b_{1nr}\right)$ is given by

$$\begin{aligned}\phi&=\frac{t_{1nr}\left(b_{1r},b_{1nr}\right)}{t_{1r}\left(b_{1r},b_{1nr}\right)}\\&=1-\frac{2\left(4+5b_{1nr}+2\left(b_{1nr}-1\right)b_{1r}+2b_{1r}^2\right)d}{\left(1+2\left(b_{1r}+b_{1nr}\right)\right)\left(e-c\right)+\left(3+b_{1r}\left(2\left(b_{1r}+b_{1nr}\right)-1\right)\right)d}.\end{aligned}$$

Note that $\phi=1$ for $d=0$ and $0<\phi<1$ for $d\in\left(0,\left(e-c\right)/22\right)$. The total tariff at the tariff vector $\mathbf{t}$ is

$$T_1=\sum_{j\in N}t_{1j}=\left(h_{1r}+\phi h_{1nr}\right)t+\left(3-b_{1r}-h_{1r}\right)t_r+\left(3-b_{1nr}-h_{1nr}\right)t_{nr}.$$

The change in the total tariff is calculated from $d\mathbf{t}$ as follows:

$$\begin{aligned}dT_1&=\left(h_{1r}+\phi h_{1nr}\right)dt+\left(3-b_{1r}-h_{1r}\right)dt_r+\left(3-b_{1nr}-h_{1nr}\right)dt_{nr}\\&=\frac{h_{1r}t_{1r}\left(b_{1r},b_{1nr}\right)+h_{1nr}t_{1nr}\left(b_{1r},b_{1nr}\right)+\left(3-b_{1r}-h_{1r}\right)\Delta t_{1r}+\left(3-b_{1nr}-h_{1nr}\right)\Delta t_{1nr}}{t_{1r}\left(b_{1r},b_{1nr}\right)}dt.\end{aligned}$$

The following notation will also be helpful:

$$\Delta T_1 = h_{1r}t_{1r}(b_{1r}, b_{1nr}) + h_{1nr}t_{1nr}(b_{1r}, b_{1nr}) + (3 - b_{1r} - h_{1r})\Delta t_{1r} + (3 - b_{1nr} - h_{1nr})\Delta t_{1nr}.$$

From (2.4) and the first-order-condition of (2.5), we have $p_i - c = x_{ij} + t_{ij} + d_{ij}$.

From (2.6), $dx_{ij} = \frac{dT_i - 7dt_{ij}}{7}$. Therefore we have:

$$\begin{aligned}
\frac{dx_{11}}{d\mathbf{t}} &= \frac{\Delta T_1}{7t_{1r}(b_{1r}, b_{1nr})};\\
\frac{dx_{1b_{1r}+1}}{d\mathbf{t}} &= \frac{\Delta T_1 - 7t_{1r}(b_{1r}, b_{1nr})}{7t_{1r}(b_{1r}, b_{1nr})};\\
\frac{dx_{1b_{1r}+h_{1r}+1}}{d\mathbf{t}} &= \frac{\Delta T_1 - 7\Delta t_{1r}}{7t_{1r}(b_{1r}, b_{1nr})};\\
\frac{dx_{1b_{1nr}+4}}{d\mathbf{t}} &= \frac{\Delta T_1 - 7t_{1nr}(b_{1r}, b_{1nr})}{7t_{1r}(b_{1r}, b_{1nr})};\\
\frac{dx_{1b_{1nr}+h_{1nr}+4}}{d\mathbf{t}} &= \frac{\Delta T_1 - 7\Delta t_{1nr}}{7t_{1r}(b_{1r}, b_{1nr})}.
\end{aligned}$$

Using these results,

$$\begin{aligned}
\frac{d}{d\mathbf{t}}\left(\hat{w}_1 + \sum_{j\in B_1'\setminus\{1\}} \pi_{1j}\right) &= \frac{d}{d\mathbf{t}}(v(X_1) - cX_1) - \frac{d}{d\mathbf{t}}\sum_{j\in N\setminus B_1'} x_{1j}^2 - d\left(\frac{d}{d\mathbf{t}}\sum_{j\in R_2} x_{1j}\right)\\
&= \sum_{j\in N}(p_1 - c)\frac{dx_{1j}}{d\mathbf{t}} - \sum_{j\in N\setminus B_1'} 2x_{1j}\frac{dx_{1j}}{d\mathbf{t}} - d\sum_{j\in R_2}\frac{dx_{1j}}{d\mathbf{t}}\\
&= \frac{1}{7t_{1r}(b_{1r}, b_{1nr})}\{h_{1r}t_{1r}(b_{1r}, b_{1nr})\Xi_1 + h_{1nr}t_{1nr}(b_{1r}, b_{1nr})\Phi_1\\
&\quad + (3 - b_{1r} - h_{1r})\Delta t_{1r}\Psi_1 + (3 - b_{1nr} - h_{1nr})\Delta t_{1nr}\Omega_1\},
\end{aligned}$$

where:

$$\begin{aligned}
\Xi_1 &= (X_1 + T_1) - 7(x_{1b_{1r}+1} + t)\\
&\quad -2(3 - b_{1r} - h_{1r})x_{1b_{1r}+h_{1r}+1} - 2(3 - b_{1nr} - h_{1nr})x_{1b_{1nr}+h_{1nr}+4};\\
\Phi_1 &= (X_1 + T_1) - 7(x_{1b_{1nr}+4} + \phi t)\\
&\quad -2(3 - b_{1r} - h_{1r})x_{1b_{1r}+h_{1r}+1} - 2(3 - b_{1nr} - h_{1nr})x_{1b_{1nr}+h_{1nr}+4};\\
\Psi_1 &= (X_1 + T_1) - 7(x_{1b_{1r}+h_{1r}+1} + t_r) + 2(4 + b_{1r} + h_{1r})x_{1b_{1r}+h_{1r}+1};\\
\Omega_1 &= (X_1 + T_1) - 7(x_{1b_{1nr}+h_{1nr}+4} + t_{nr}) + 2(4 + b_{1nr} + h_{1nr})x_{1b_{1nr}+h_{1nr}+4}.
\end{aligned}$$

The proof that $\frac{d}{d\mathbf{t}}\left(\hat{w}_1 + \sum_{j\in B_1'\setminus\{1\}} \pi_{ij}\right) < 0$ proceeds in two steps. First we show that, at $\mathbf{t}(b_{1r}', b_{1nr}')$, it is the case that $\frac{d}{d\mathbf{t}}\left(\hat{w}_1 + \sum_{j\in B_1'\setminus\{1\}} \pi_{ij}\right) < 0$. Second, we show that $\frac{d^2}{d\mathbf{t}^2}\left(\hat{w}_1 + \sum_{j\in B_1'\setminus\{1\}} \pi_{ij}\right) < 0$.

Step 1. At $\mathbf{t}\left(b_{1r}',b_{1nr}'\right)$, the optimal tariffs $t_{1r}\left(b_{1r}',b_{1nr}'\right)$ and $t_{1nr}\left(b_{1r}',b_{1nr}'\right)$ are chosen to satisfy $\Psi_1 = 0$ and $\Omega_1 = 0$ respectively. (Note that Ψ_1 and Ω_1 are the derivatives of $\hat{w}_1 + \sum_{j\in B_1'\setminus\{1\}}\pi_{ij}$ with respect to t_{1r} and t_{1nr} respectively; $t_{1r}\left(b_{1r}',b_{1nr}'\right)$ and $t_{1nr}\left(b_{1r}',b_{1nr}'\right)$ are the optimal tariffs of the size $b_{1r}'+b_{1nr}'$ TA on $3-b_{1r}'$ regional non-members and $3-b_{1nr}$ non-regional non-members respectively, given free trade among the $b_{1r}'+b_{1nr}'$ members.) It remains to show that, at $\mathbf{t}\left(b_{1r}',b_{1nr}'\right)$, the terms Ξ_1 and Φ_1 are both strictly negative. (Of course, due to oligopoly distortions, Ξ_1 and Φ_1 could only be zero if trade subsidies were allowed).

At $\mathbf{t}\left(b_{1r}',b_{1nr}'\right)$, $x_{11} =, ..., = x_{1b_{1r}+h_{1r}}$, $x_{14} =, ..., x_{1b_{1nr}+h_{1nr}+3}$ (unless $b_{1nr} = h_{1nr} = 0$, in which case $x_{14} =, ..., x_{1b_{1nr}+h_{1nr}+4}$), and $t = 0$. Also,

$$\begin{aligned}
X_1 &= b_{1r}'x_{11} + \left(3-b_{1r}'\right)x_{1b_{1r}+h_{1r}+1} + b_{1nr}'x_{1b_{1nr}+h_{1nr}+3} + \left(3-b_{1nr}'\right)x_{1b_{1nr}+h_{1nr}+4};\\
T_1 &= \left(3-b_{1r}'\right)t_{1r}\left(b_{1r}',b_{1nr}'\right) + \left(3-b_{1nr}'\right)t_{1nr}\left(b_{1r}',b_{1nr}'\right).
\end{aligned}$$

Then we have

$$\begin{aligned}
\Xi_1 &= -4x_{11} + b_{1nr}'x_{1b_{1nr}+h_{1nr}+3}\\
&\quad -\left(3-b_{1r}'\right)\left(x_{11} + x_{1b_{1r}+h_{1r}+1} - t_{1r}\left(b_{1r}',b_{1nr}'\right)\right)\\
&\quad -\left(3-b_{1nr}'\right)\left(x_{1b_{1nr}+h_{1nr}+4} - t_{1nr}\left(b_{1r}',b_{1nr}'\right)\right).
\end{aligned}$$

Now, observing that $x_{11} = x_{1rm}$, $x_{1b_{1r}+h_{1r}+1} = x_{1rnm}$, $x_{1b_{1nr}+h_{1nr}+3} = x_{1nrm}$ and $x_{1b_{1nr}+h_{1nr}+4} = x_{1nrnm}$, we can use (2.7)-(2.10) to substitute for x_{11}, $x_{1b_{1r}+h_{1r}+1}$, $x_{1b_{1nr}+h_{1nr}+3}$ and $x_{1b_{1nr}+h_{1nr}+4}$, which obtains

$$\Xi_1 = -\frac{7\left(2\left(e-c\right) + \left(3 + b_{inr}\left(3 + 2\left(b_{ir} + b_{inr}\right)\right)\right)d\right)}{10 + b_{ir}\left(5 + 2b_{ir}\right) + b_{inr}\left(4 + b_{ir}\right) + b_{inr}^2} < 0.$$

Next observe that, after simplification,

$$\begin{aligned}
\Phi_1 &= -4x_{1b_{1r}+h_{1r}+3} + b_{1r}'x_{11}\\
&\quad -\left(3-b_{1nr}'\right)\left(x_{1b_{1r}+h_{1r}+3} + x_{1b_{1nr}+h_{1nr}+4} - t_{1nr}\left(b_{1r}',b_{1nr}'\right)\right)\\
&\quad -\left(3-b_{1r}'\right)\left(x_{1b_{1r}+h_{1r}+1} - t_{1r}\left(b_{1r}',b_{1nr}'\right)\right).
\end{aligned}$$

Adopting the same basic approach used to simplify Ξ_1, we then have

$$\Phi_1 = -\frac{7\left(2\left(e-c\right) - \left(5 + b_{inr}\left(3 + 2\left(b_{ir} + b_{inr}\right)\right)\right)d\right)}{10 + b_{ir}\left(5 + 2b_{ir}\right) + b_{inr}\left(4 + b_{ir}\right) + b_{inr}^2}$$

We can see straight away that for $d = 0$ it is the case that $\Phi_1 < 0$, and that Φ_1 is increasing in d. We then find by substitution that for $d = (e-c)/22$, $b_{ir} = 3$ and $b_{inr} = 2$, it is the case that $\Phi_1 = 0$. It follows immediately that $\Phi_1 < 0$ for all $b_{1r} \in \{1,2,3\}$ and $b_{1nr} \in \{0,1,2\}$ and $d \in [0,(e-c)/22)$.

Step 2. We can write the second order condition directly as

$$\begin{aligned}\frac{d^2}{d\mathbf{t}^2}\left(\hat{w}_1 + \sum_{j\in B_1'\backslash\{1\}} \pi_{ij}\right) \quad = \quad & \frac{1}{(7t_{1r})^2}\Big(-\left(3-b_{1r}'\right)\left(35+15b_{1r}'+2\left(b_{1r}'\right)^2\right)\Delta t_{1r}^2 \\ & +\left(3-b_{1r}'\right)\left(15+4b_{1r}'+2b_{1nr}'\right)\left(h_{1r}t_{1r}+h_{1nr}t_{1nr}\right)\Delta t_{1r} \\ & -\left(3-b_{1nr}'\right)\left(-14+15b_{1nr}'+2\left(b_{1nr}'\right)^2\right)\Delta t_{1nr}^2 \\ & +\left(3-b_{1nr}'\right)\left(15+4b_{1nr}'+2b_{1r}'\right)\left(h_{1r}t_{1r}+h_{1nr}t_{1nr}\right)\Delta t_{1nr} \\ & -2\left(3-b_{1r}'-b_{1nr}'\right)\left(h_{1r}t_{1r}'+h_{1nr}t_{1nr}'\right)^2 \\ & -\left(3-b_{1nr}'\right)\left(7+8b_{1r}'-2b_{1nr}'\left(3-b_{1r}'\right)+2\left(b_{1r}'\right)^2\right)\Delta t_{1r}\Delta t_{1nr}.\end{aligned}$$

Using the functions for $t_{ir}(b_{ir}, b_{inr})$, $t_{ir}(b_{ir}', b_{inr}')$, $t_{inr}(b_{ir}, b_{inr})$ and $t_{inr}(b_{ir}', b_{inr}')$, substitution reveals that the second order condition is negative for all feasible values $b_{1r}' \in \{1,2,3\}$ and $b_{1nr}' \in \{0,1,2,3\}$, given $d \in [0,(e-c)/22)$.

Proof of Proposition 5. Without loss of generality, take country 1 as an example. (The cases for all other countries are analogous.) Write down two welfare functions for country 1: $w_1\{\{1,2,3\},\{4,5,6\}\}$ and $w_1\{\{1,4,5,6\},\{2,3\}\}$. The first measures the welfare of country 1 when it is in a regional TA and all countries in the other region are in a second regional TA. The second welfare function measures welfare when country 1 joins a TA with the countries in the other region while countries 2 and 3 form a TA. To calculate $w_1\{\{1,2,3\},\{4,5,6\}\}$, note that country 1 sets a tariff $t_{inr}^*(3,0)$ on all imports from the other region, and country 1's exports also face $t_{inr}^*(3,0)$ from all countries in the other region. Trade within regions is free. Using these tariffs in (2.7)-(2.10) and substituting the resulting expressions into (2.11), we obtain

$$w_1\{\{1,2,3\},\{4,5,6\}\} = \frac{3\left(387(e-c)^2 - 134(e-c)d + 1072d^2\right)}{2450}$$

For $w_1\{\{1,4,5,6\},\{2,3\}\}$, country 1 sets $t_{ir}^*(1,3)$ on imports from non-members in its own region. country 1's exports face tariffs $t_{ir}^*(2,0)$ from non-members in its own region. Trade between country 1 and the countries in the other region is free. Using these tariffs in (2.7)-

(2.10), and substituting the resulting expressions into (2.11), we obtain

$$w_1\left\{\{1,4,5,6\},\{2,3\}\right\} = \frac{3\left(26442\,(e-c)^2 - 44336\,(e-c)\,d + 92225d^2\right)}{163592}.$$

We can now see that

$$\begin{aligned} w_1\left\{\{1,4,5,6\},\{2,3\}\right\} &> w_1\left\{\{1,2,3\},\{4,5,6\}\right\} \text{ for } d=0; \\ w_1\left\{\{1,4,5,6\},\{2,3\}\right\} &< w_1\left\{\{1,2,3\},\{4,5,6\}\right\} \text{ for } d=(e-c)/22. \end{aligned}$$

We can also see that both $w_1\left\{\{1,2,3\},\{4,5,6\}\right\}$ and $w_1\left\{\{1,4,5,6\},\{2,3\}\right\}$ are decreasing in d for $d \in (0,(e-c)/22)$ but $w_1\left\{\{1,4,5,6\},\{2,3\}\right\}$ is decreasing more rapidly. So we can find a unique value of $d \in (0,(e-c)/22)$, called d', at which $w_1\left\{\{1,2,3\},\{4,5,6\}\right\} = w_1\left\{\{1,4,5,6\},\{2,3\}\right\}$;

$$d' = \frac{3\left(7225156 - 385\sqrt{338226178}\right))}{25290313}(e-c) \simeq 0.017\,(e-c)$$

□

Proof of Proposition 6. By Proposition 2, member welfare of a given TA is decreasing in the size of each of the other TAs that exist. Therefore, the highest feasible level of welfare is achieved when a country is a member of a TA and all non-members of its TA are singletons.

It remains to establish the TA structure that maximizes member welfare (given that all non-members are singletons). The result is seen clearly if we take each case in turn, starting with the smallest possible TA and increasing its size while evaluating member welfare at each point. First, it follows from Proposition 3 that if two singletons form a two-member TA this must increase member welfare. We now establish that if both members are in the same region this yields a higher level of welfare than if each member is in a different region. Without loss of generality, assume that country 1 forms a 2-country TA either with country 2 in its own region or with country 4 in the other region. Welfare would be $w_1\left\{\{1,2\},\{3\},\{4\},\{5\},\{6\}\right\}$ or $w_1\left\{\{1,4\},\{2\},\{3\},\{5\},\{6\}\right\}$ respectively. To calculate $w_1\left\{\{1,2\},\{3\},\{4\},\{5\},\{6\}\right\}$, note that country 1 levies a tariff $t^*_{ir}(2,0)$ and $t^*_{inr}(2,0)$ on imports from regional and non-regional non-members respectively. The non-member from R_1 levies a tariff $t^*_{ir}(1,0)$ on imports from country 1, and non-members from R_2 levy a tariff $t^*_{inr}(1,0)$ on imports from country 1. Substituting these tariffs into (2.7)-(2.10) and substituting appropriately into (2.11) yields

$$w_1\left\{\{1,2\},\{3\},\{4\},\{5\},\{6\}\right\} = \frac{889\,(e-c)^2 - 999\,(e-c)\,d + 2205d^2}{1859}.$$

To calculate $w_1\{\{1,4\},\{2\},\{3\},\{5\},\{6\}\}$, note that country 1 levies a tariff $t^*_{ir}(1,1)$ and $t^*_{inr}(1,1)$ on imports from regional and non-regional non-members respectively. The non-members from R_1 levy $t^*_{ir}(1,0)$ on imports from country 1, and non-members from R_2 levy $t^*_{inr}(1,0)$ on imports from country 1. Substituting these tariffs into (2.7)-(2.10) and substituting appropriately into (2.11) yields

$$w_1\{\{1,4\},\{2\},\{3\},\{5\},\{6\}\} = \frac{7112\,(e-c)^2 - 4404\,(e-c)\,d + 16431d^2}{14872}.$$

Welfare under the two TA configurations is equal for $d=0$ and the latter yields a lower level of welfare for $d>0$, with the difference increasing in the size of d.

The same basic approach can be used to establish that the 3-member TA that maximizes a member's welfare is where all members are in the same region, and that a 3-member regional TA yields a higher level of per-member welfare than a 2-member regional TA:

$$w_1(\{1,2,3\},\{4\},\{5\},\{6\}) = \frac{5787\,(e-c)^2 - 3114\,(e-c)\,d + 13362d^2}{11830}.$$

We can also calculate the level of welfare of country 1 if a non-regional member is included; $w_1\{\{1,2,3,4\},\{5\},\{6\}\}$. In that case, country 1 imposes a tariff $t^*_{inr}(3,1)$ on imports from non-members, and non-members impose a tariff $t^*_{inr}(1,0)$ on imports from country 1. Substituting these tariffs into (2.7)-(2.10), and making the appropriate substitution into (2.11), we have

$$w_1\{\{1,2,3,4\},\{5\},\{6\}\} = \frac{333\,(e-c)^2 - 262\,(e-c)\,d + 915d^2}{676}.$$

We can now see that

$$\begin{aligned} w_1\{\{1,2,3,4\},\{5\},\{6\}\} &> w_1\{\{1,2,3\},\{4\},\{5\},\{6\}\} \text{ for } d=0 \\ w_1\{\{1,2,3,4\},\{5\},\{6\}\} &< w_1\{\{1,2,3\},\{4\},\{5\},\{6\}\} \text{ for } d=(e-c)/22. \end{aligned}$$

We can also see that $w_1\{\{1,2,3,4\},\{5\},\{6\}\}$ is declining in d for $d \in (0,(e-c)/22)$. So we can find a unique value of $d \in (0,(e-c)/22)$, called d'', at which $w_1\{\{1,2,3\},\{4\},\{5\},\{6\}\} = w_1\{\{1,2,3,4\},\{5\},\{6\}\}$;

$$d'' = \frac{1471 - 2\sqrt{433615}}{5301}(e-c) \simeq 0.029\,(e-c).$$

Finally, we must check that a 5-member TA does not yield a higher level of welfare than either a 4-member TA or a 3-member TA. As for all previous cases, a member obtains a higher payoff if all countries in its own region are members of the TA. Thus

$$w_1 \{\{1,2,3,4,5\},\{6\}\} = \frac{12145\,(e-c)^2 - 11262\,(e-c)\,d + 37450d^2}{24674}.$$

Since, for $d = 0$,

$$w_1 \{\{1,2,3,4\},\{5,6\}\} > w_1 \{\{1,2,3,4,5\},\{6\}\},$$

and since $w_1 \{\{1,2,3,4,5\},\{6\}\}$ has a steeper negative slope in d than $w_1 \{\{1,2,3,4\},\{5,6\}\}$, it follows that, for all $d \in [0,(e-c)/22)$,

$$w_1 \{\{1,2,3,4\},\{5\},\{6\}\} > w_1 \{\{1,2,3,4,5\},\{6\}\}.$$

Similar calculations show that free trade yields a lower level of per-member welfare than $w_1 \{\{1,2,3,4\},\{5\},\{6\}\}$ and $w_1 \{\{1,2,3\},\{4\},\{5\},\{6\}\}$. $\square$

References

[1] Aghion, P., P. Antras and E. Helpman (2007); "Negotiating Free Trade." *Journal of International Economics* 73: 1-30.

[2] Arnold, T. and M. Wooders (2005); "Dynamic Club Formation with Coordination." Vanderbilt University Department of Economics Working Paper No. 05-W22.

[3] Bagwell, K. and R. Staiger (1997a); "Multilateral Tariff Cooperation During the Formation of Regional Free Trade Areas." *International Economic Review*, 38: 291-319.

[4] Bagwell, K. and R. Staiger (1997b); "Multilateral Tariff Cooperation During the Formation of Customs Unions." *Journal of International Economics*, 42: 91-123.

[5] Bagwell, K. and R. Staiger (1998); "Will Preferential Arrangements Undermine the Multilateral Trading System?" *Economic Journal* 108(449): 1162-1182.

[6] Baier, S.L. and J.H. Bergstrand (2004); "Economic Determinants of Free Trade Agreements." *Journal of International Economics*, 64: 29-63.

[7] Baldwin, R.E. (1996); "A Domino Theory of Regionalism." In Baldwin, R.E., P.J. Haaparanta and J. Kiander (eds.) *Expanding the Membership of the EU*, Cambridge University Press, Cambridge.

[8] Baldwin, R.E. (2009); "Big-Think Regionalism: A Critical Survey." Published in Antoni Estevadeordal, Kati Suominen, Robert Teh (eds), *Regional Rules in the Global Trading System*, Cambridge University Press; 1st edition (June 30, 2009). Also see NBER WP 14056.

[9] Bergstrand, J.H., P. Egger and M. Larch (2009); "Economic Determinants of the Timing of Preferential Trade Agreement Formations and Enlargements." University of Notre Dame Typescript.

[10] Bhagwati, J. (1993); "Regionalism and Multilateralism: An Overview." In De Melo, J. and A. Panagariya (eds.) *New Dimensions in Regional Integration*, Cambridge University Press, Cambridge.

[11] Bhagwati, J., D. Greenaway and A. Panagariya (1998); "Trading Preferentially: Theory and Policy." *Economic Journal*, 108: 1128-1148.

[12] Bloch, F. (1996); "Sequential Formation of Coalitions in Games with Externalities and Fixed Payoff Division." *Games and Economic Behavior*, 14: 90-123.

[13] Bloch, F. (2003); "Non-cooperative Models of Coalition Formation in Games with Spillovers." Chapter 2 in C. Carraro (ed) 2003, *The Endogenous Formation of Economic Coalitions* published by Edward Elgar, Cheltenham, UK and Northampton, USA.

[14] Bond, E. (2001); Multilateralism vs. Regionalism: Tariff Cooperation and Interregional Transport Costs." *Regionalism and Multilateralism*, Sajal Lahiri, ed. Routledge Press, 2001.

[15] Bond, R., R. Riezman and C. Syropoulos (2004); "A Strategic and Welfare Theoretic Analysis of Free Trade Areas." *Journal of International Economics*, 64: 1-27.

[16] Bond, E. and C. Syropoulos (1996); "Trading Blocs and the Sustainability of Inter-Regional Cooperation." Published in M. Conzoneri, W.J. Ethier and V. Grilli (eds.) *The New Transatlantic Economy*, London: Cambridge University Press.

[17] Bond, E., C. Syropoulos and L.A. Winters (2001); "Deepening of Regional Integration and Multilateral Trade Agreements." *Journal of International Economics*, 53: 335-361.

[18] Brander, J. and B. Spencer (1984) “Tariff protection and imperfect competition.” chapter in H. Kierkowski (ed.) *Monopolistic Competition and International Trade*, Oxford University Press, Oxford.

[19] Egger, P. and M. Larch (2008); “Interdependent Preferential Trade Agreement Memberships: An Empirical Analysis.” *Journal of International Economics* 76(2): 384–399.

[20] Ethier, W. (1998); “Regionalism in a Multilateral World.” *Journal of Political Economy*, 106(6): 1214-1245.

[21] Goto, J., and K. Hamada (1999); “Regional Economic Integration and Article XXIV of the GATT.” *Review of International Economics* 7(4): 555-570.

[22] Grossman, G.M. and E. Helpman (1995); “Trade Wars and Trade Talks.” *Journal of Political Economy*, 103(4): 675-708.

[23] Hart, S. and M. Kurz (1983); “Endogenous Formation of Coalitions.” *Econometrica*, 51(4): 1047-1064.

[24] Krishna, P. (1998); “Regionalism and Multilateralism: A Political Economy Approach.” *Quarterly Journal of Economics*, 113(1): 227-251.

[25] Maggi, G. and A. Rodriguez-Clare (1998); “The Value of Trade Agreements in the Presence of Political Pressures.” *Journal of Political Economy*, 106: 574-601.

[26] McClaren, J. (2002); “A Theory of Insidious Regionalism.” *Quarterly Journal of Economics* 117: 571– 608.

[27] Mrazova, M., D. Vines and B. Zissimos (2009); “Is the WTO’s Article XXIV Bad?” CEPR Discussion Paper no. 7144.

[28] Ornelas, E. (2005a); “Endogenous Free Trade Agreements and the Multilateral Trading System.” *Journal of International Economics*, 67: 471-497.

[29] Ornelas, E. (2005b); “Trade Creating Free Trade Areas and the Undermining of Multilateralism.” *European Economic Review*, 49: 1717–1735.

[30] Riezman (1985); “Customs Unions and the Core.” *Journal of International Economics*, 19: 355-66.

[31] Riezman, R. (1999); "Can Bilateral Trade Agreements Help to Induce Free Trade?" *Canadian Journal of Economics*, 32(3): 751-766.

[32] Saggi, K. (2006); "Preferential Trade Agreements and Multilateral Tariff Cooperation." *International Economic Review*, 47(1): 29-57.

[33] Seidmann, D. (2009); "Preferential Trading Arrangements as Strategic Positioning." *Journal of International Economics*, 79(1): 143-159.

[34] Syropoulos, C. (1999); "Customs Unions and Comparative Advantage." *Oxford Economic Papers*, 51(2): 239-266.

[35] Syropoulos, C. (2002); "Tariff Preferences in Customs Unions." *Economic Journal* 112(481): 625-48.

[36] Viner, J. (1950); *The Customs Union Issue.* (Washington, DC: Carnegie Endowment for International Peace).

[37] Whalley, J. (1985); *Trade Liberalization among Major World Trading Areas.*, MIT Press, Cambridge, Mass.

[38] Yi, S.-S. (1996); "Endogenous Formation of Customs Unions under Imperfect Competition: Open Regionalism is Good." *Journal of International Economics*, 41: 151-175.

[39] Yi, S.-S. (2003); "Endogenous Formation of Economic Coalitions: A Survey of the Partition Function Approach." Chapter 2 in C. Carraro (ed) 2003, *The Endogenous Formation of Economic Coalitions* published by Edward Elgar, Cheltemham, UK and Northampton, USA.