Should the WTO require free trade agreements to eliminate internal tariffs?

Kamal Saggi, Woan Foong Wong, Halis Murat Yildiz

PII: S0022-1996(18)30418-5
DOI: doi:10.1016/j.jinteco.2018.10.003
Reference: INEC 3181

To appear in: *Journal of International Economics*

Received date: 23 September 2017
Revised date: 15 October 2018
Accepted date: 16 October 2018

Please cite this article as: Kamal Saggi, Woan Foong Wong, Halis Murat Yildiz , Should the WTO require free trade agreements to eliminate internal tariffs?. Inec (2018), doi:10.1016/j.jinteco.2018.10.003

Should the WTO require free trade agreements to eliminate internal tariffs?*

Kamal Saggi[†], Woan Foong Wong[‡], Halis Murat Yildiz[§]

*We are grateful to James Amegashie, Kyle Bagwell, Kamran Bilir, James Lake, Alex Maynard, Ralph Ossa, Robert Staiger, Andrey Stoyanov, and two anonymous referees for their constructive suggestions. We also thank the participants at the 2016 UECE Lisbon Meetings on Game Theory and Applications, 2017 Midwest Economic Theory and International Trade Conference at Southern Methodist University, 2017 Southern Economic Association Meetings, 2018 ASSA meetings, and 2018 InsTED Workshop at Syracuse University as well as the seminar participants at the University of Winnipeg for helpful comments.

[†]Department of Economics, Vanderbilt University, Nashville, TN 37235-1828. E-mail: kamal.saggi@vanderbilt.edu.

[‡]Department of Economics, University of Oregon, Eugene, Oregon, 97403. E-mail: wfwong@uoregon.edu.

[§]Department of Economics, Ryerson University, 350 Victoria Street, Toronto, ON, Canada M5B 2K3. E-mail: hyildiz@ryerson.ca.

Abstract

In a three-country model of endogenous free trade agreements (FTAs), we study the effects of requiring FTA members to eliminate tariffs on one another, as is essentially stipulated under current WTO rules. We explain why, in the absence of such a requirement, FTAs members impose positive tariffs on each other even when maximizing their joint welfare. We show that requiring FTA members to eliminate internal tariffs induces them to *lower* their external tariffs. Such external trade liberalization by FTA members undermines the prospects of global free trade since it *reduces* the non-member's incentive to enter into trade agreements with them.

Keywords: Free Trade Agreements, Tariffs, Customs Unions, World Trade Organization, Coalition proof Nash equilibrium, Welfare.

JEL Classifications: F11, F12.

1 Introduction

Under the current rules of the World Trade Organization (WTO), countries entering into a preferential trade agreement (PTA) are required to eliminate tariffs on " substantially all trade" with each other.This condition and other related provisions governing PTAs are specified in Article XXIV of the General Agreement on Tariffs and Trade (GATT), the key multilateral agreement governing international trade in goods amongst WTO members. This paper develops a model of endogenous trade agreements to investigate the welfare implications of this *free internal trade requirement* facing PTAs at the WTO as well as the effect it has on the likelihood of achieving global free trade. In the existing literature, Article XXIV has often been invoked as a justification for the *assumption* that PTA members impose zero tariffs on each other. Though reasonable, this approach masks the incentives underlying the tariff-setting behavior of PTA members and, by design, fails to shed light on the consequences of requiring them to fully liberalize internal trade.

We focus on free trade agreements (FTAs), the most commonly occurring type of PTA in today's global economy. Our conceptual approach to the formation of trade agreements follows Saggi and Yildiz (2010) who develop an equilibrium theory of FTAs in a modified version of the three-country competing exporters framework of Bagwell and Staiger (1999a).[1] Assuming FTA members impose zero tariffs on one another, they compare the relative merits of bilateralism and multilateralism as alternative routes to global trade liberalization. Although the WTO system sanctions discrimination in the specific form of PTAs, it also requires all member countries to grant most favored nation (MFN) status to one another which generally forbids discrimination on their part. Thus, we begin with a *WTO-consistent benchmark* scenario under which FTA members are required to eliminate tariffs on each other and the non-member is obligated to follow the MFN principle of non-discrimination when setting its tariffs on FTA members. We compare this WTO-consistent benchmark with a scenario of *unconstrained preferential liberalization* wherein FTA members have the freedom to implement jointly optimal internal tariffs as opposed to having to eliminate them as a precondition for forming the FTA.[2]

A comparison of the WTO-consistent scenario with the unconstrained preferential liberalization scenario delivers several interesting results. First, we show that if FTA members choose internal tariffs to maximize their joint welfare, they indeed have an incentive to impose positive tariffs on one another. The intuition for this surprising result rests on the interplay between two mechanisms: the *lack of external tariff coordination* between FTA members and the *complementarity of imports tariffs*. Since FTA members set their external tariffs independently, each member fails to take into account the benefits that its external tariff

[1] Saggi et. al (2013) build on Saggi and Yildiz (2010) by considering trade agreements that take the form of customs unions as opposed to FTAs.

[2] While GATT Article XXIV requires FTA members to impose zero internal tariffs on each other, FTA members do not always abide by this restriction. An analysis of PTAs involving 85 countries and 90 percent of world trade in 2007 found that roughly two-thirds of tariff lines with MFN rates greater than 15 percent were not reduced through PTAs (Bagwell et. al, 2016 and WTO, 2011). Our model sheds light on the consequences of such non-compliance on the part of PTA members regarding the free internal trade requirement of GATT Article XXIV.

confers on its partner – if an FTA member raises its tariff on the non-member, exports of its FTA partner to its market increase while those of the non-member decrease. Thus, because each FTA member ignores the impact of its external tariff on the welfare of its partner, the individually optimal external tariffs of FTA members are too low from the perspective of maximizing their joint welfare.

The existence of tariff complementarity and the lack of external tariff coordination together imply that, while coordinating their internal tariffs, FTA members deliberately choose to set positive internal tariffs on each other: doing so commits each of them to a higher external tariff on the non-member country thereby bringing their individually optimal external tariffs closer to jointly optimal ones. To confirm the role that external tariff coordination plays in generating positive internal tariffs within an FTA, we consider a setting where FTA members can coordinate their external as well as internal tariffs, as they might be able to do under a customs union (CU). Under such a case, members indeed find it optimal to engage in free internal trade. This result suggests that the free internal trade requirement of Article XXIV is likely to be more binding for FTAs relative to CUs.[3] Although there is some evidence that FTAs tend to have more excluded sectors than CUs, there is a lack of comprehensive empirical evidence on internal tariffs and excluded sectors in FTAs and CUs.[4] Freund and Ornelas (2010) highlight the wide range of implementation rates of PTAs as a vital research area that has received little attention.[5]

The second major insight delivered by our analysis is that requiring FTA members to eliminate internal tariffs *benefits* the non-member since it leads to *lower* external tariffs on the part of FTA members. This result, driven by tariff complementarity, is noteworthy since part of the original intent behind the design of Article XXIV may have been to minimize any potential negative effects of FTAs on non-member countries. Ostensibly, this objective was met by prohibiting FTA members from raising their external tariffs on outsiders. However, in our model, FTA members have no incentive to increase their external tariffs on the non-member country anyway.[6] Thus, the Article XXIV stipulation that FTA members cannot raise tariffs on outsiders may actually do little to protect the interests of outsiders. The idea that the requirement of free internal trade amongst FTA members could imply *lower tariffs for outsiders*

[3] This result is in line with Kennan and Riezman (1990), Yi (1996), Bagwell and Staiger (1998), Cadot, de Melo, and Olarreaga (1999), Freund (2000), and Ornelas (2007).

[4] Exception includes Liu (2010) which studies how the influence of special interest groups relative to voters affects the choice between partial-scope (formed under the Enabling Clause of GATT) and full-fledged trade agreements.

[5] Using product exclusions from 15 FTAs signed by the US, EU, Japan, and Canada, Damuri (2012) shows that 7 percent of tariff lines are excluded, either temporarily or permanently. Agriculture and food products are the most protected products while manufactured products are the least protected. These product exclusions are also different across FTAs with different partners, highlighting the discriminatory feature of FTAs. Product exclusion is correlated with the regime of trade protection proxied by MFN tariff rates. Studying the bilateral trade agreements of countries in ASEAN, APEC, and South Asia, Menon (2009) also finds that the most commonly excluded sector is agriculture. In the example of Japan's trade agreement with Mexico, 13 percent of Mexico's exports to Japan are excluded from the trade agreement. In comparison, CUs like the European Union are fully implemented (Freund and Ornelas, 2010) while Mercusor only excluded the sugar and automobile industries (Olarreaga and Soloaga, 1998).

[6] This result is not specific to models with tariff complementarity but is also present in models with endogenous protection (Richardson (1993, 1995)) as well as models with firm-delocation externalities (Suwanprasert (2017)). Richardson (1993) shows that since FTAs will shift imports away from non-member countries, FTA countries have an incentive to lower external tariffs to shift these imports back if the diverted imports reduce its welfare. Suwanprasert (2017) augments Ossa (2011) by allowing for all countries to trade with each other and finds that whenever countries 1 and 2 agree on a bilateral trade agreement, country 3 always gains from the agreement although it is not involved in the negotiations. The firms in country 3 benefit from gaining better access to country 1's market even though they face more competition from manufacturing firms in country 2. Bond, Riezman, and Syropoulos (2004) finds that, at constant rest of the world tariffs, countries that join free trade agreements reduce their external tariffs on outside countries. They present their results as stronger than Bagwell and Staiger (1998)'s tariff complementarity findings since the external tariff fall is so large that it improves the rest of the world's terms of trade.

was probably unforeseen at the time the relevant GATT rules were crafted. Instead, it seems more likely that the requirement of zero internal tariffs was designed to promote trade creation amongst FTA members. Our analysis demonstrates that, somewhat surprisingly, it is the Article XXIV requirement of free internal trade within an FTA that ends up protecting the non-member as opposed to the restriction imposed on the external tariffs of FTAs.

Our third major result pertaining to the free internal trade requirement of Article XXIV is that having such a requirement makes it *harder* to achieve global free trade. The logic for this result is as follows. By lowering the external tariffs of FTA members, the free internal trade requirement of Article XXIV makes it *less* attractive for the non-member to enter into trade agreements with them – by staying out, it remains free to impose its optimal import tariffs while facing relatively lower tariffs in the markets of FTA countries due to the disciplining force of the free internal trade requirement.[7] Thus, the free internal trade requirement of Article XXIV might facilitate some degree of *free-riding* in the WTO system by allowing non-member countries to benefit from reductions in external tariffs of FTA members (that result from their internal trade liberalization) *without* having to offer any tariff cuts of their own. Thus, our overall message is somewhat nuanced: when circumstances are such that achieving global free trade is not possible, the free internal trade requirement of Article XXIV increases world welfare by lowering tariffs world-wide but, at the same time, it also reduces the likelihood of reaching global free trade.

In Section 5 of the paper we show that our results are robust to two alternative tariff setting scenarios and to a fairly wide range of endowment asymmetry across countries. First, we relax the assumption that countries seeking to form FTAs set their MFN tariffs non-cooperatively since WTO members do seem to cooperate in the setting of their MFN tariffs even though such cooperation is hardly perfect. To this end, we allow countries to engage in a limited degree of cooperation by assuming that they assign some weight to the welfare of other countries while setting their MFN tariffs. We show that our main results regarding the impact of the free internal trade requirement continue to hold even when countries do not set their tariffs in a fully non-cooperative manner. In our second robustness exercise, we indirectly address the issue of the *extent of enforceability* of the free internal trade provision of Article XXIV. We do this by examining a scenario where Article XXIV imposes a ceiling on the internal tariffs of an FTA. Under such a scenario, we show that the free riding incentive continues to be the pivotal force in determining the prospects of global free trade: the tighter the ceiling imposed on the internal tariffs of FTAs (i.e. the more it binds), the lower the external tariffs of FTA members. A lower ceiling brings us closer to the free internal trade scenario, making it *less* attractive for the non-member to enter into trade agreements with FTA members which in turn undermines global free trade. Finally, we demonstrate that our main results extend to the case when all three countries are asymmetric, unlike our benchmark case wherein two countries are fully symmetric.

Since Bhagwati (1991), the literature has paid significant attention to whether PTAs serve as *building* or *stumbling* blocs for multilateral trade liberalization. Early theoretical research on this issue generally took PTAs to be exogenously given and focused on how PTA

[7] The the free-rider problem caused by MFN during multilateral trade negotiations has been examined by Johnson (1965), Caplin and Krishna (1988), and Ludema and Mayda (2009, 2013). Wong (2017) shows that the free rider problem removes global free trade as a stable outcome in multilateral trade negotiations.

membership affects the incentives that countries have for participating in multilateral trade liberalization (see, for example, Krishna, 1998; Ornelas, 2005a, 2005b). More recent studies, such as Goyal and Joshi (2006), Aghion et al. (2007), Furusawa and Konishi (2007), and Seidman (2009) consider endogenous PTAs but ignore the possibility of trade liberalization on an MFN basis. Under this approach, PTAs are seen as building blocs so long as their pursuit eventually leads to global free trade. However, Saggi and Yildiz (2010), Saggi et. al (2013), Missios et al. (2016) and Stoyanov and Yildiz (2015) have argued that PTAs ought to be seen as building blocks only if the freedom to pursue PTAs (granted to WTO members by GATT Article XXIV) is *necessary* for achieving global free trade. An attractive feature of this line of research is that it treats both preferential and multilateral liberalization as being endogenous. This paper follows this approach and furthers the literature on the building versus stumbling bloc question by showing that the free internal trade requirement of Article XXIV makes it harder to achieve global free trade, i.e., it reduces the likelihood that PTAs act as building blocs.

The paper proceeds as follows. In the next section, we introduce our competing exporters model of trade between three countries. Section 3 outlines the main policy scenarios we investigate: the WTO-consistent scenario where FTA members are required to engage in free internal trade and the unrestricted preferential liberalization scenario where members are free to impose non-zero internal tariffs on one another. In section 4, we solve for the equilibrium outcome for both of these policy scenarios and compare their differences. Section 5 confirms the robustness of our results via three important extensions and Section 6 concludes.

2 Tariffs and trade

Our underlying trade model is an adapted version of the partial equilibrium " competing exporters" framework developed by Bagwell and Staiger (1999a) to analyze the effects of PTAs. There are three asymmetrically endowed countries: $i, j, \text{and } k$ and three (non-numeraire) goods: $I, J, \text{and } K$.[8] Each country's market is served by two competing exporters and I denotes the good that corresponds to the upper case value of i. Country i is endowed with zero units of good I and e_i units of the other two goods.

The demand for good z in country i is given by

$$d(p_i^z) = \alpha - p_i^z \text{ where } z = I, J, \text{ or } K \quad (1)$$

As is well known, the above demand functions can be derived from a utility function of the form $U(c^z) = u(c^z) + w$ where c^z denotes consumption of good z; w denotes the numeraire good; and $u(c^z)$ is quadratic and additively separable in each of the three goods. Country i must import good I in order to consume it and can import it from either trading partners j or k.

Let t_{ij} be the tariff imposed by country i on its imports of good I from country j. Ruling out prohibitive tariffs yields the following no-arbitrage conditions:

$$p_i^I = p_j^I + t_{ij} = p_k^I + t_{ik} \quad (2)$$

Let m_i^I be country i's imports of good I. Since country i has no endowment of good I, we have

[8] All countries have large enough endowments of the freely traded numeraire good that they consume in positive quantities.

$$m_i^I = d(p_i^I) = \alpha - p_i^I \quad (3)$$

Each country's exports of a good must equal its endowment of that good minus its local consumption:

$$x_j^I = e_j - [\alpha - p_j^I] \quad (4)$$

Market clearing for good I requires that country i's imports equal the total exports of the other two countries:

$$m_i^I = \sum_{j \neq i} x_j^I \quad (5)$$

Equations (2) through (5) imply that the equilibrium price of good I in country i equals:

$$p_i^I = \frac{1}{3}\left(3\alpha - \sum_{j \neq i} e_j + \sum_{j \neq i} t_{ij}\right) \quad (6)$$

A country's terms of trade motive for import tariffs is evident from equation (6): only a third of a given increase in either of its tariffs is passed on to domestic consumers in the form of a price increase, with the rest of the burden falling on the shoulders of foreign exporters.

From a welfare perspective, given the partial equilibrium nature of the model, it suffices to consider only protected goods. A country's welfare is defined as the sum of consumer surplus, producer surplus, and tariff revenue over all such goods:

$$w_i = \sum_z CS_i^z + \sum_z PS_i^z + TR_i \quad (7)$$

Using equations (2) through (6) one can easily obtain welfare of country i as a function of endowment levels and tariffs. Let aggregate world welfare be defined as the sum of each country's welfare:

$$ww = \sum_i w_i. \quad (8)$$

Before proceeding further, we note that in order to guarantee non-negative exports and positive tariffs under all trade policy regimes in all scenarios, we impose the following parameter restriction throughout the paper on the country endowment sizes: $\max\{e_i, e_j, e_k\} \leq \frac{5}{4}\min\{e_i, e_j, e_k\}$.[9]

Suppose countries do not enter into any type of trade agreement with each other. Then, in accordance with MFN clause, country i must set the same non-discriminatory tariff on both its partners, $t_{ij} = t_{ik}$. Let t_i^M denote country i's optimal MFN tariff where

$$t_i^M \equiv \arg\max w_i(t_{ij}, t_{ik}) \text{ such that } t_{ij} = t_{ik} \quad (9)$$

Now let us consider how the formation of an FTA between two countries, say i and j, affects the non-member country. It is useful to begin with exogenously given internal and external tariffs and consider how variations in these tariffs affect the non-member. Let the pair of internal tariffs set by FTA members i and j on each other be denoted by (τ_{ij}, τ_{ji}). Our first point is simply that, all else equal, the non-member's welfare declines if the internal tariffs within the FTA decline (we call this as the *discrimination effect)*:

$$\frac{\partial w_k}{\partial \tau_{ij}} > 0 \text{ and } \frac{\partial w_k}{\partial \tau_{ji}} > 0 \quad (10)$$

This is due to the competing exporter framework where j and k are competing for i's

[9] Calculations supporting this restriction and all of the results reported in the paper are contained in Appendix subsection 7.1.

market. As such, a decrease in τ_{ij} means that FTA partner j has more market access to i relative to non-member k which lowers k's welfare.

Consider now the relationship between the internal and external tariffs of an FTA between countries i and j. We assume that FTA members first choose their internal tariffs (τ_{ij}, τ_{ji}) to maximize their joint welfare and then, given the internal tariffs, each FTA member independently chooses its external tariff to maximize its own welfare. Thus, as a member of a bilateral FTA with country j, country i chooses t_{ik} to $\max w_i(t_{ik}; \tau_{ij})$.[10] The optimal external tariff of FTA member i as a function of its internal tariff on FTA member j is given by

$$t_{ik}^*(\tau_{ij}) \equiv \arg\max_{t_{ik}} w_i(t_{ik}; \tau_{ij})$$

Using the first order condition for the above problem, we can show the following:

$$\frac{dt_{ik}^*(\tau_{ij})}{d\tau_{ij}} > 0 \tag{11}$$

i.e. the individually optimal external tariff of an FTA member country is increasing in its internal tariff on the other member country. In other words, there is *tariff complementarity* between the internal and external tariffs of FTA member countries. This tariff complementarity implies that the deeper the degree of internal trade liberalization in an FTA, the lower the tariffs that FTA members impose on the non-member.[11]

The above tariff analysis shows that the preferential trade liberalization undertaken by FTA members has two conflicting effects on the non-member country. On one hand, the non-member loses from the discrimination that is inherent to FTAs (equation (10)). On the other hand, the internal liberalization within an FTA induces each member to lower its tariff on the non-member (equation (11)). Furthermore, when external tariffs are chosen by FTA members to maximize their respective welfare, the tariff complementarity effect *outweighs* the discrimination effect so that the larger the degree of internal trade liberalization between FTA members, the higher the non-member's welfare, i.e., at $t_{ik} = t_{ik}^*(\tau_{ij})$ and $t_{jk} = t_{jk}^*(\tau_{ji})$ we have:

$$\frac{\partial w_k}{\partial \tau_{ij}} < 0 \text{ and } \frac{\partial w_k}{\partial \tau_{ji}} < 0$$

Now consider tariff setting within an FTA. While setting their internal tariffs, FTA members jointly solve

$$\max_{\tau_{ij}, \tau_{ji}} \left[w_i(\tau_{ij}, \tau_{ji}, t_{ik}^*(\tau_{ij}), t_{jk}^*(\tau_{ji})) + w_j(\tau_{ij}, \tau_{ji}, t_{ik}^*(\tau_{ij}), t_{jk}^*(\tau_{ji})) \right]$$

In other words, while setting their internal tariffs, FTA member account for the fact that each of them chooses an individually optimal external tariff. The first order condition for τ_{ij} is given by

$$\frac{\partial w_i}{\partial \tau_{ij}} + \frac{\partial w_i}{\partial t_{ik}} \frac{dt_{ik}^*(\tau_{ij})}{d\tau_{ij}} + \frac{\partial w_j}{\partial \tau_{ij}} + \frac{\partial w_j}{\partial t_{ik}} \frac{dt_{ik}^*(\tau_{ij})}{d\tau_{ij}} = 0$$

which can be rewritten as

---

[10] Due to the structure of the model, a country's individually optimal tariffs are independent of the tariffs of its trading partners (since these apply to different goods). In other words, country i's choice of t_{ik} only depends upon t_{ij} and is independent of all other tariffs.

[11] This result extends beyond the present framework and can be found in models with endogenous protection (Richardson (1993, 1995)) as well as models with firm-delocation externalities (Suwanprasert (2017)).

$$\frac{\partial(w_i+w_j)}{\partial\tau_{ij}}+\frac{dt_{ik}^*}{d\tau_{ij}}\left[\frac{\partial(w_i+w_j)}{\partial t_{ik}}\right]=0 \quad (12)$$

Note that

$$\frac{\partial(w_i+w_j)}{\partial\tau_{ij}}<0$$

i.e., all else equal, an increase in country i's internal tariff lowers the joint welfare of FTA members but, as noted above in (11), due to tariff complementarity we have $\frac{dt_{ik}^*}{d\tau_{ij}}>0$. Furthermore, at the individually optimal external tariff chosen by country i the following must hold:

$$\frac{\partial w_i}{\partial t_{ik}}=0$$

But since $\frac{\partial w_j}{\partial t_{ik}}>0$, it immediately follows from (12) that at the individually optimal external tariff chosen by country i we must have

$$\frac{\partial(w_i+w_j)}{\partial t_{ik}}>0$$

Intuitively, since country i does not take into account the effect of its tariff on its partner country, it is jointly welfare improving for FTA members to raise their external tariffs above their individually optimal tariffs. As a result, though positive internal tariffs hurts FTA members by lowering internal trade, they also benefit them by committing them to higher external tariffs on the non-member. As a result, FTA members find it jointly optimal to impose positive internal tariffs on each other. Let the optimal internal tariffs set by countries i and j on each other be denoted by $(\tau_{ij}^*,\ \tau_{ji}^*)$.

We summarize the key messages of the above analysis in the following lemma:

Lemma 1: *(i) The larger the degree of internal trade liberalization undertaken by FTA members, the higher the welfare of the non-member country and (ii) FTA members impose strictly positive internal tariffs on each other, i.e.* $\tau_{ij}^*>0$ and $\tau_{ji}^*>0$.

As noted above, the first result is due to the tariff complementarity effect of an FTA dominating its discrimination effect. The intuition behind part (*ii*) is more subtle: due to the lack of external tariff coordination in an FTA, each FTA member does not take into account the fact that an increase in its external tariff benefits its FTA partner since its exports compete with those of the non-member. Thus, the individually optimal external tariffs of FTA members are too low from the perspective of maximizing their joint welfare. The coordination of internal tariffs prior to the independent setting of external tariffs, provides FTA members with a partial remedy to this problem. Due to the tariff complementarity effect, deliberately setting positive internal tariffs on each other *commits* FTA members to imposing higher external tariffs on the non-member country, thereby bringing their individually optimal external tariffs closer to jointly optimal ones.

The intuition underlying the tariff complementarity between external and internal tariffs of an FTA is quite robust and clean. As Maggi (2014) notes, if two countries possessing market power sign an FTA, they start to import more from each other and less from non-members and this trade diversion reduces their incentives to manipulate their terms of trade vis-a-vis

non-members, which ultimately results in lower external tariffs on their part. Tariff complementarity arises in a variety of different models of international trade including oligopoly models of intra-industry trade (Ornelas (2005a), Saggi (2004), Saggi and Yildiz (2011) and Stoyanov and Yildiz (2015)), general equilibrium Ricardian models (Kennan and Riezman (1990) and Bond et al. (2004)), and competitive partial equilibrium models with integrated markets (Bagwell and Staiger (1999a,b), Saggi and Yildiz (2010), Saggi et al. (2013)).[12] Specifically, Bond et al. (2004) show that, at constant tariffs in the rest of the world, countries that join free trade agreements reduce their external tariffs on outsiders. This is a stronger result than Bagwell and Staiger (1998)'s tariff complementarity findings since the fall in external tariffs of member countries is so large that it improves the rest of the world's terms of trade. Empirical support for this type of tariff complementarity has been provided by Bohara et al. (2004), Estevadeordal et al. (2008), Calvo-Pardo et al, (2009), and Mai and Stoyanov (2015). Using the data and approach of Estevadeordal et al. (2008), Crivelli (2016) shows that the strength of the tariff complementarity effect depends on the initial tariff levels.

To confirm the role that tariff coordination plays in generating positive internal tariffs within an FTA, suppose FTA members could coordinate their internal *and* external tariffs, as they might be able to do under a customs union (CU). Then, both members jointly solve the following maximization problem[13]

$$\max_{\tau_{ij},\tau_{ji},t_{ik},t_{jk}} \left[w_i(\tau_{ij},\tau_{ji},t_{ik},t_{jk}) + w_j(\tau_{ij},\tau_{ji},t_{ik},t_{jk})\right]$$

Since tariffs of different countries in our framework apply to different goods, it suffices to focus on country i's choices of τ_{ij} and t_{ik}. Differentiating the objective function with respect to τ_{ij} we have

$$\frac{\partial(w_i + w_j)}{\partial \tau_{ij}} < 0$$

If members coordinate their external tariffs, an FTA becomes equivalent to a CU in our model and members find it optimal to engage in free internal trade since their joint welfare is strictly decreasing in each of the internal tariffs. The optimal external tariff of the CU between i and j (t_{ik}^{u}) is defined, following the above, by $\frac{\partial(w_i+w_j)}{\partial t_{ik}} = 0$. It is straightforward to show that CU members impose higher external tariffs than FTA members: $t_{zk}^{u} > t_{zk}^{*}$ where $z = i, j$. Thus, due to the dual coordination of internal and external tariffs, a CU between two countries yields (*i*) deeper internal trade liberalization and (*ii*) higher external tariffs relative to an FTA between them.[14]

3 Endogenous trade agreements

The two policy scenarios that we study are formalized as follows:

(*a*) *WTO-consistent scenario*: This scenario is captured by a three stage game of trade liberalization under which countries abide by both Article I and Article XXIV of GATT. In the first

[12] It is important to note that all of these models rely on specific quasi-linear or Cobb-Douglas preferences. In order to understand whether tariff complementarity holds under general conditions, under oligopoly model, Saggi and Yildiz (2009) isolate sufficiency conditions under which a PTA is less likely to impose a positive external tariff relative to that under MFN.

[13] When both external and internal tariffs are coordinated, the tariff problem compresses to a single stage.

[14] Mrázová, Vines, and Zissimos (2013) study Article XXIV's constraint on coordinated external tariff increase and its impact on CU formation.

stage, countries enter into FTAs with one another (the process of FTA formation is described in greater detail below). In the second stage, given the trade policy regime that results from the first stage, countries choose their optimal tariffs. If an FTA is formed, its members practice free internal trade while imposing individually optimal external tariffs on the non-member who, in accordance with MFN, imposes non-discriminatory tariffs on the two member countries. At the third stage of the game, given trade agreements and tariffs, international trade and consumption take place.

(*b*) *Unconstrained preferential liberalization scenario*: This scenario is formalized as a four stage game that proceeds as follows. The first stage of the game remains the same as the first stage of the WTO-consistent scenario. At the second stage, given the policy regime, FTA members set their internal tariffs to maximize their joint welfare. As opposed to the WTO-consistent scenario described in (*a*), the internal tariffs of an FTA do not have to be reduced to zero. Next, all countries independently and simultaneously choose their external tariffs. At the last stage of the game, international trade and consumption occur.

We now describe the process of FTA formation that occurs during the first stage of the game and is common to both scenarios.

The process of FTA formation: At the first stage of the game, each country announces whether or not it wants to sign an FTA with each of the other two countries. Denote country i's announcement by σ_i and its strategy set by S_i:

$$S_i = \{\{\phi,\phi\},\{j,\phi\},\{\phi,k\},\{j,k\}\} \quad (13)$$

where $\{\phi,\phi\}$ denotes an announcement in favor of no FTAs, $\{j,\phi\}$ an announcement in favor of an FTA with only country j; $\{\phi,k\}$ in favor of an FTA with only country k; and $\{j,k\}$ in favor of FTAs with both of them. Since a trade agreement requires consent from both sides, we posit the following mapping between various announcements profiles and the types of trade agreements that countries can form:

(*i*) No two announcements match or the only matching announcements are $\{\phi,\phi\}$. All of these announcement profiles yield no agreement $\langle\Phi\rangle$. Under the WTO consistent and unconstrained preferential liberalization scenarios, all countries impose their optimal MFN tariffs on one another.

(*ii*) Two countries announce each others' name and there is no other matching announcement: i.e., $j \in \sigma_i$ and $i \in \sigma_j$ while $i \notin \sigma_k$ and/or $k \notin \sigma_i$ and $j \notin \sigma_k$ and/or $k \notin \sigma_j$. All of these announcements yield an FTA between countries i and j denoted by $\langle ij\rangle$ under which members eliminate internal tariffs under the WTO consistent scenario while imposing their jointly optimal internal tariffs under the unconstrained preferential liberalization scenario. Under both scenarios, members impose their individually optimal external tariffs on the non-member k.

(*iii*) Country i announces in favor of signing an FTA with countries j and k while countries j and/or k announce only in favor of signing an FTA with country i: i.e. $\sigma_i = \{j,k\}$; $i \in \sigma_j$; and $i \in \sigma_k$ while $k \notin \sigma_j$ and/or $j \notin \sigma_k$. This set of announcements yields a pair of independent FTAs (i.e. a hub and spoke trading regime) with i as the common member denoted by $\langle ij,ik\rangle$ (or simply $\langle ih\rangle$). Under a hub and spoke agreement $\langle ih\rangle$, hub country i sets zero tariffs (optimal under both scenarios) on exports from the spoke countries while the spokes solve the same tariff problems as they do under a bilateral FTA with country i.

(*iv*) All countries announce each others' names, i.e., the announcement profile is

$\Omega^F \equiv \{\sigma_i = \{j,k\}, \sigma_j = \{i,k\}, \sigma_k = \{i,j\}\}$. This announcement profile yields the global free trade regime $\langle F \rangle$.

Note that since an FTA between two countries can arise only if it is mutually acceptable to both sides, multiple announcement profiles can map into the same agreement. For example, the FTA $\langle ij \rangle$ can result from the following announcement profiles. First, when countries i and j call only each other, *regardless* of the nature of country k's announcement: if $\sigma_i = \{j, \phi\}$ and $\sigma_j = \{i, \phi\}$, then $\langle ij \rangle$ is the outcome for all four possible announcements on the part of country k, i.e., for $\sigma_k = \{\phi, \phi\}$, $\{i, \phi\}$, $\{\phi, j\}$ and $\{i, j\}$. Note that country k's announcement has no bearing upon the outcome when neither of the other two countries' announce its name. Second, when countries i and j announce each other's names and either one or both of them also announce country k but country k does not reciprocate, i.e. all of the following types of announcements map into the FTA $\langle ij \rangle$: (a) $\sigma_i = \{j,k\}$ and $\sigma_j = \{i, \phi\}$ but $i \notin \sigma_k$ or (b) $\sigma_i = \{j, \phi\}$ and $\sigma_j = \{i,k\}$ but $j \notin \sigma_k$ or (c) $\sigma_i = \{j,k\}$ and $\sigma_j = \{i,k\}$ but $\sigma_k = \{\phi, \phi\}$.

When analyzing the above games, we only consider those Nash equilibria that are *coalition-proof*. Following Bernheim et al. (1987): " ... an agreement is coalition-proof if and only if it is Pareto efficient within the class of self-enforcing agreements. In turn, an agreement is self-enforcing if and only if no proper subset (coalition) of players, taking the actions of its complement as fixed, can agree to deviate in a way that makes all of its members better off." Therefore, a coalition proof Nash equilibrium (CPNE) is a Nash equilibrium that is immune to all *self-enforcing* coalitional deviations.

4 Equilibrium agreements

In order to simplify our exposition, we make the following assumption:

Assumption 1: *Countries l and l' are larger importers than country s*: $e_s = \theta e \geq e_l = e_{l'} = e$ *where* $1 \leq \theta \leq 5/4$.[15]

It is worth pointing out that, in our model, all countries can affect their terms of trade via import tariffs. Although country s has a weaker ability to manipulate its terms of trade relative to the other two, it is not a " small" country in the traditional sense of the term wherein it would be a price-taker on world markets.

Recall that each country's endowment of the (unique) good it imports is zero and that asymmetry in endowments translates directly into asymmetries in export volumes. In other words, an increase in a country's endowment in this model increases its exports of non-numeraire/ protected goods without increasing its imports of such goods (since the model is partial equilibrium in nature and lacks any income effects). Indeed, since the country with the largest endowment of non-numeraire goods faces relatively smaller suppliers, its imports of such goods are smaller. Therefore, from here on, country s is called the " smaller importing country" and l and l' the " larger importing countries." Note that the smaller importing

[15] The qualitative nature of our results is robust to a scenario where all three countries are asymmetric, such as when $e_s = \theta_s e \geq e_m = \theta_m e \geq e_l = e$ where $\frac{5}{4} \geq \theta_s \geq \theta_m \geq 1$. But since the key insights can be illustrated more easily in the simpler case where the two larger countries are symmetric, we first proceed with this assumption. Section 5 extends this baseline model to case of three asymmetric countries and shows that our main results continue to hold.

country is a relatively larger exporter and, in a non-cooperative equilibrium, it faces higher tariffs compared to the larger importers.

We proceed as follows. First, we study FTA formation in our WTO-consistent benchmark scenario, where FTA members are forced to eliminate internal tariffs, and show that no two countries have an incentive to form a bilateral trade agreement in order to exclude the third country. Instead, it is the strength of the free-riding incentive of the non-member country that proves pivotal in determining whether or not global free trade emerges as the equilibrium outcome. Next, we derive the equilibrium trade agreements under the unrestricted preferential liberalization scenario where FTA members are free to impose positive internal tariffs on each other. In equilibrium, FTA members utilize this freedom and they also end up imposing higher external tariffs relative to the WTO-consistent benchmark. This in turn reduces the free-riding incentive of the non-member country and therefore furthers the cause of global free trade by making it more attractive for it to enter into trade agreements with the other two countries. On the other hand, when global free trade is infeasible, the free internal trade requirement of Article XXIV raises global welfare by lowering internal *and* external tariffs of FTA countries.

4.1 WTO-consistent benchmark

In this section, we derive the equilibrium trade agreements under our benchmark scenario where FTA members engage in free internal trade and the non-member country follows MFN. Let country i's welfare as a function of the underlying trade policy regime r be denoted by $w_i(r)$, where $r = \langle\Phi\rangle, \langle ij\rangle, \langle ih\rangle$, or $\langle F\rangle$ and it is understood that all countries impose optimal tariffs consistent with regime r. For example, if $r = \langle ij\rangle$ then countries i and j eliminate internal tariffs on each other respectively while imposing the tariffs t_{ik}^* and t_{jk}^* on country k. Let $\Delta w_i(r-v)$ denote the difference between country i's welfare under trade agreements r and v: $\Delta w_i(r-v) \equiv w_i(r) - w_i(v)$, where $r, v = \langle\Phi\rangle, \langle ij\rangle, \langle ih\rangle$, or $\langle F\rangle$. Furthermore, let $\theta_i(r-v)$ denote the critical threshold of asymmetry at which country i is indifferent between regimes r and v.

We first state the following lemma that explains how differences in market power across countries lead them to have asymmetric preferences over various trade regimes:

Lemma 2: *In the WTO-consistent approach to the formation of trade agreements, the following holds:*

(i) Each country prefers to form a bilateral FTA with the larger importer relative to the smaller one: $\Delta w_l(ll' - sl) > 0$ *for all* θ.

(ii) The smaller importer (s) *has an incentive to form an additional bilateral FTA under any trade regime except for when it is a non-member facing an FTA between the other two countries.*

(iii) Each larger importer prefers being a non-member under a bilateral FTA to being a spoke under a hub and spoke regime while the smaller importer does so only when the degree of endowment asymmetry is sufficiently small: $\Delta w_{l'}(lh - sl) < 0$ and $\Delta w_{l'}(sh - sl) < 0$ *for all* θ *and* $\Delta w_s(lh - ll') < 0$ *when* $\theta < \theta_s(lh - ll')$.

(iv) All countries prefer being the hub under a hub and spoke regime relative to all other trade policy regimes: $\Delta w_i(ih - \Phi) > 0$; $\Delta w_i(ih - F) > 0$ *and* $\Delta w_i(ih - ij) > 0$ *for all*

$i, j = s, l, l'$ *and* $i \neq j$.

Part (*i*) of Lemma 1 follows from two reinforcing effects. The larger a country's trading partner's import volume, the larger the increase in export surplus it enjoys from the elimination of its partner's optimal tariff and the smaller the loss it suffers from its own trade liberalization since its tariff reduction applies to a smaller volume of imports. Thus, a country prefers to form a bilateral FTA with the larger importer amongst its two trading partners. The second part of Lemma 1 states that the smaller importer (i.e. country s) has an incentive to form an additional FTA under any given regime except when the existing regime is $\langle ll' \rangle$ and the endowment asymmetry is sufficiently large (see part (*iii*)). This implies that, generally speaking, choices of the larger importing countries are critical in determining whether or not an FTA between two asymmetric countries arises. Finally, part (*iv*) says that being a hub country is better for all countries irrespective of their size relative to all other trade policy regimes. Note in particular that, relative to free trade, the hub country enjoys privileged access to both spoke countries while its domestic surplus is no different. Moreover, this privileged access in export markets is so desirable that a hub country has no incentive to unilaterally revoke any of its FTAs.

While FTA members discriminate against the non-member, we know from the above tariff analysis that the internal trade liberalization of an FTA actually benefits the non-member. This raises the possibility that, starting from no agreement $\langle \Phi \rangle$, the formation of an FTA makes all countries better off (i.e. is Pareto improving relative to $\langle \Phi \rangle$). Indeed, we can show that the smaller country benefits from the formation of an FTA between large countries only when the degree of endowment asymmetry is sufficiently small:

$$\Delta w_s(ll' - \Phi) > 0 \text{ when } \theta < \theta_s(ll' - \Phi) \tag{14}$$

Second, while the larger non-member (country l') always benefits from the formation of $\langle sl \rangle$, the larger member country benefits from the formation of $\langle sl \rangle$ only when the degree of asymmetry is sufficiently small:

$$\Delta w_{l'}(sl - \Phi) > 0 \text{ when } \theta < \theta_l(sl - \Phi) \tag{15}$$

Therefore, we find the following:

Proposition 1: *Relative to no agreement* $\langle \Phi \rangle$ *wherein all countries impose their optimal Nash tariffs on each other, the FTA* $\langle ll' \rangle$ *is Pareto-improving iff* $\theta < \theta_s(ll' - \Phi)$ *while the the FTA* $\langle sl \rangle$ *is Pareto-improving iff* $\theta < \theta_l(sl - \Phi)$.

Armed with the underlying incentives identified by Lemma 2, we are now ready to determine the CPNE of the WTO-consistent game of trade agreements. We proceed by considering each of the announcement profiles that yield the various trade policy regimes in turn. First, consider the announcement profile leading to global free trade $\langle F \rangle$. First note from part (*ii*) of the Lemma 2 that smaller importer (i.e. country s) has no incentive to participate in any deviation (unilateral or coalitional). Thus, if there exists a coalitional deviation, it must involve countries l and l'. Taking country s' announcement fixed at $\{l, l'\}$, countries l and l' have an incentive to jointly deviate from their respective announcements $\{s, l'\}$ and $\{s, l\}$ to $\{\phi, l'\}$ and $\{\phi, l\}$ in order to exclude country s from a free trade network when country s is a sufficiently small importer:

$$\Delta w_l(F - ll') < 0 \text{ when } \theta > \theta_l(F - ll') \quad (16)$$

The above result establishes the existence of an *exclusion incentive*: when the endowment asymmetry is sufficiently pronounced (i.e. $\theta > \theta_l(F - ll')$) the two larger importers prefer a bilateral FTA between themselves to global free trade. Furthermore, since world welfare is higher under free trade than under a bilateral FTA, it follows that the non-member country is better off under free trade relative to the bilateral FTA $\langle ij \rangle$.

Is the joint exclusion incentive of the two larger importers self-enforcing? The answer to this key question is in the negative. To see why, suppose each country announces in favor of an FTA with both its trading partners. Starting with these announcements the two larger importers have an incentive to exclude the smaller country by jointly altering their announcements from Ω^F (which yields free trade) to $\Omega_1^{ll'} = \{\sigma_l = \{\phi, l'\}, \sigma_{l'} = \{\phi, l\}, \sigma_s = \{l, l'\}\}$ thereby altering the associated trade regime from free trade to the bilateral FTA $\langle ll' \rangle$. However, from part (*iv*) of Lemma 2 we know that each country's most preferred trading arrangement is a hub and spoke regime with itself serving as the hub. It follows then that, holding constant the announcement of the excluded country at $\sigma_s = \{l, l'\}$, each member of the deviating coalition (l or l') has an incentive to alter its announcement to include country *s*. For example, country l has an incentive to alter its announcement from $\sigma_l = \{\phi, l'\}$ to $\sigma_l = \{s, l'\}$ which alters the trade regime from $\langle ll' \rangle$ to $\langle lh \rangle$. Since the welfare of a hub is higher than that of a member country in a single FTA – see part (*iv*) of Lemma 2 – the original coalitional deviation of countries l and l' from Ω^F to $\Omega_1^{ll'}$ is *not* self-enforcing. Thus, in a nutshell, the lure of a hub and spoke trading arrangement makes any joint deviation from Ω^F to an announcement profile that supports a bilateral FTA between any two countries not-self enforcing.

Consider now announcement deviations that convert the trade regime from $\langle F \rangle$ to $\langle \Phi \rangle$. It is easy to see that since all countries are better off under free trade relative to $\langle \Phi \rangle$, no two countries have an incentive to deviate from Ω^F to an announcement profile that yields $\langle \Phi \rangle$. For example, holding $\sigma_s = \{l, l'\}$, countries l and l' have no incentive to jointly deviate from their respective announcements $\{s, l'\}$ and $\{s, l\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$. Based on the above discussion, the only possible type of self-enforcing deviation from Ω^F that we need to consider is a unilateral deviation from Ω^F by one of the large importers. To this end, we find that there exists no incentive of a large country (say l) to unilaterally deviate from its announcements $\{s, l'\}$ to any announcement that leads to a hub and spoke regime under which country s is a hub and itself a spoke:

$$\Delta w_l(F - sh) = \Delta w_{l'}(F - sh) \geq 0 \text{ for all } \theta \quad (17)$$

Then two unilateral deviation incentives remain to be examined: (*i*) country l unilaterally deviating from $\{s, l'\}$ to $\{\phi, l'\}$:

$$\Delta w_l(F - l'h) = \Delta w_{l'}(F - lh) < 0 \text{ when } \theta > \theta_l(F - l'h) \quad (18)$$

and (*ii*) country l unilaterally deviating from $\{s, l'\}$ to $\{\phi, \phi\}$:

$$\Delta w_l(F - sl') = \Delta w_{l'}(F - sl) < 0 \text{ when } \theta > \theta_l(F - sl') \quad (19)$$

We find that $\theta_l(F - sl') < \theta_l(F - l'h)$ and thus the announcement profile leading to $\langle F \rangle$ is CPNE whenever $\theta \leq \theta_l(F - sl')$.

What if $\langle F \rangle$ is not a CPNE, as is the case when the degree of country asymmetry is sufficiently large ($\theta > \theta_l(F - sl')$)? We can quickly rule out the various announcement profiles leading to the hub and spoke regimes as candidates for CPNE. To see why, recall from part (*iii*)

of Lemma 2 that a larger spoke country (say l) under $\langle sh \rangle$ and $\langle l'h \rangle$ has an incentive to unilaterally deviate from its respective announcements $\{s, \phi\}$ and $\{\phi, l'\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$, leading to a deviation from $\langle sh \rangle$ to $\langle sl' \rangle$ and from $\langle l'h \rangle$ to $\langle sl' \rangle$. Since these unilateral deviations are self-enforcing, any announcement profile leading to a hub and spoke regime cannot be a CPNE.

Next, we consider the various announcement profiles that lead to no agreement $\langle \Phi \rangle$. Since countries l and l' have an incentive to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, \phi\}$ to $\{\phi, l'\}$ and $\{\phi, l\}$ in order to form $\langle ll' \rangle$, this joint deviation is self-enforcing. As a result, any announcement profile that yields $\langle \Phi \rangle$ cannot be a CPNE.

The only remaining candidates for CPNE are the announcement profiles that lead to bilateral FTAs. We start with those profiles that yield an FTA between the smaller importer and one of the larger ones, say $\langle sl \rangle$. We find that, when $\theta > \theta_l(sl - \Phi)$, country l has an incentive to unilaterally deviate from its announcement $\{s, \phi\}$ to $\{\phi, \phi\}$ thereby converting the trade policy regime from the bilateral FTA $\langle sl \rangle$ to no agreement $\langle \Phi \rangle$. Second, we know from part (*iv*) of Lemma 2 that the coalitional announcement deviation that converts $\langle sl \rangle$ to $\langle ll' \rangle$ is *not* self-enforcing since the common member country (i.e. country l) has an incentive to further deviate to become the hub country, taking the announcement of its partners as fixed. Third, from the discussion above, the coalitional announcement deviation that replaces $\langle sl \rangle$ by $\langle F \rangle$ is self-enforcing only when $\theta \leq \theta_l(F - sl')$. Finally, it is immediate from part (*iii*) of Lemma 2 that country l' has no incentive to engage in any coalitional announcement deviations that replace $\langle sl \rangle$ by $\langle sh \rangle$ or $\langle sl \rangle$ by $\langle lh \rangle$. As a result, the announcement profile leading to $\langle sl \rangle$ is a CPNE whenever $\theta_l(F - sl') \leq \theta \leq \theta_l(sl - \Phi)$.

Finally, we consider the bilateral FTA between the two larger countries, i.e., $\langle ll' \rangle$. First, as before, the coalitional announcement deviation from $\langle ll' \rangle$ to $\langle F \rangle$ occurs $\theta \leq \theta_l(F - ll')$ and it is self-enforcing when $\theta \leq \theta_l(F - sl')$. Second, we can show that when $\theta > \theta_s(lh - ll')$, country s and either of the larger countries (say l) have an incentive to jointly deviate from their respective announcements $\{\phi, \phi\}$ and $\{\phi, l'\}$ to $\{l, \phi\}$ and $\{s, l'\}$, leading to a deviation from $\langle ll' \rangle$ to $\langle lh \rangle$ and this deviation is self-enforcing. Since $\theta_s(lh - ll') < \theta_l(F - sl')$, these self-enforcing announcement deviations cover the entire parameter space and thus the announcement profile supporting $\langle ll' \rangle$ is not a CPNE.

We summarize the main findings of the above analysis below:

Proposition 2: *The equilibria of the WTO-consistent game of trade liberalization where FTA members are required to practice free internal trade and the non-member to follow MFN are as follows:*

(i) Free trade $\langle F \rangle$ *is the equilibrium agreement when* $\theta \leq \theta_l(F - sl')$.[16]

(ii) An asymmetric bilateral FTA $\langle sl \rangle$ *(or* $\langle sl' \rangle$*) is the equilibrium when* $\theta_l(F - sl') \leq \theta \leq \theta_l(sl - \Phi)$.

(iii) There exists no equilibrium if $\theta > \theta_l(sl - \Phi)$.

[16] We should note here that, technically speaking, the equilibrium is the announcement profile Ω^F that yields free trade as the agreement. In what follows, for expositional ease, we state our results directly in terms of various trade agreements that emerge as equilibrium outcomes as opposed to the announcement profiles that support them.

The above proposition relates the degree of underlying asymmetry to the nature of equilibrium agreements. Part (*i*) simply says that if the degree of endowment asymmetry is sufficiently small, free trade is the equilibrium outcome. It is important to reiterate that while the exclusion incentives of larger importing countries go unexercised in equilibrium, each large importing country's incentive to unilaterally deviate from free trade proves critical for determining the viability of free trade. Part (*ii*) states that if the degree of endowment asymmetry is sufficiently large, only an asymmetric FTA ($\langle sl \rangle$ or $\langle sl' \rangle$) is the equilibrium – in such a situation, one of the larger importing countries prefers being a non-member to participating in any bilateral or multilateral agreements. Note from the above discussion that the bilateral FTA between the two larger countries $\langle ll' \rangle$ fails to arise in equilibrium. Finally, part (*iii*) of Proposition 1 says that there exists no CPNE if the degree of endowment asymmetry is very large. In such a situation, our theory offers no guidance regarding which of the trade regimes should be expected to arise in equilibrium.[17]

What if Article XXIV allows FTAs to set positive internal tariffs? Next we allow this possibility.

4.2 Unconstrained preferential liberalization

Here, we consider the scenario of *unconstrained preferential liberalization* wherein FTA member countries jointly choose their internal tariffs before independently setting their external tariffs. Recall that, due to the existence of tariff complementarity in our model, the deeper the internal trade liberalization in an FTA, the lower the external tariffs of member countries. As a result, when allowed, member countries set positive internal tariffs on each other and this incomplete internal trade liberalization means that the degree of tariff complementarity here is smaller relative to the WTO-consistent benchmark case.

Under a hub and spoke agreement $\langle ih \rangle$, hub country i has a trade agreement with both countries j and k and its internal tariffs are chosen to maximize the joint welfare of all three countries which leads to zero internal tariffs: $\tau_{ij}^*(ih) = \tau_{ik}^*(ih) = 0$, while the spoke countries' tariffs solve the same problem as they do under a bilateral trade agreement so that $t_{jk}^*(ih) = t_{jk}^*(ij)$.

Let country i's welfare as a function of the underlying trade agreement r with positive internal tariffs be denoted by $w_i(\hat{r})$ and let $\Delta w_i(\hat{r} - \hat{v})$ denote the difference between country i's welfare under trade agreements r and v with positive internal tariffs: $\Delta w_i(\hat{r} - \hat{v}) \equiv w_i(\hat{r}) - w_i(\hat{v})$. The following lemma explains the preferences of asymmetric countries over trade regimes when member countries are able to impose internal tariffs before setting their external tariffs:

Lemma 3: *When member countries of an FTA choose their internal tariffs jointly before*

[17] When we compare this parameter space under different scenarios, we do not take any stand regarding the trade regimes that can arise.

setting their individually optimal external tariffs, the following holds:

(i) Starting from no agreement $\langle\Phi\rangle$*, all countries have an incentive to form a bilateral FTA:* $\Delta w_i(\widehat{ij} - \Phi) > 0$ *for all* θ *and* $i, j = s, l, l'$.

(ii) A large importer prefers a bilateral FTA with the other larger importer relative to the smaller one: $\Delta w_l(\widehat{ll'} - \widehat{sl}) > 0$ *for all* θ.

(iii) The smaller importer has an incentive to form a bilateral FTA under any trade regime.

(iv) Each larger importer prefers being a non-member under a bilateral FTA to being a spoke under a hub and spoke regime provided endowments are sufficiently asymmetric across countries: $\Delta w_l(\widehat{l'h} - \widehat{sl'}) < 0$ *when* $\theta > \theta_l(\widehat{l'h} - \widehat{sl'})$ *and* $\Delta w_l(\widehat{sh} - \widehat{sl'}) < 0$ *when* $\theta > \theta_l(\widehat{sh} - \widehat{sl'})$.

(v) All countries prefer being the hub country under a hub and spoke regime relative to no agreement as well as to being a member under a bilateral FTA: $\Delta w_i(\widehat{ih} - \Phi) > 0$ *and* $\Delta w_i(\widehat{ih} - \widehat{ij}) > 0$ *for all* $i, j = s, l, l'$ *and* $i \neq j$.

The intuition behind part (*i*) of Lemma 3 is that when member countries under a bilateral FTA can coordinate internal tariffs before setting their individually optimal external tariffs, they *partially internalize* the effects of their external tariffs on one another and this increases the incentive of larger importing countries to form a bilateral FTA. We find that, relative to the WTO-consistent benchmark case, the incentives for forming FTAs are generally stronger under unconstrained preferential liberalization since FTA members are less constrained and can therefore achieve higher levels of welfare under FTAs. Furthermore, due to the joint determination of internal tariffs, a country's preference to form a bilateral FTA with the larger of its two trading partners is even stronger. Parts (*iii*) and (*iv*) of Lemma 3 differ from part (*ii*) and part (*iii*) of Lemma 2 in an important way: while coordinating their internal tariffs, FTA members deliberately choose to set positive internal tariffs. Doing so leads each member to impose a higher external tariff on the non-member country relative to our WTO-consistent benchmark case. This in turn decreases the incentive of the non-member to stay outside the FTA, whether it faces a bilateral FTA or finds itself as a spoke under a hub and spoke regime. Finally, part (*v*) of Lemma 3 says that being a hub country is better for all countries (irrespective of their size) relative to no agreement and to being a member of a bilateral FTA.

An interesting question is whether bilateral FTA formation is more or less likely to be Pareto-improving over no agreement when FTA members are free to impose positive internal tariffs on each other. Since tariff complementarity is weaker when FTA members are not constrained by Article XXIV, the non-member country's relative situation is worse under the unconstrained liberalization scenario relative to the WTO-consistent benchmark scenario. As indicated above, since member countries always benefit from forming an FTA relative to no agreement, the Pareto-improvement condition of a bilateral FTA with internal tariffs relies only on the welfare of the non-member country. We first find that, starting from no agreement, a larger country always benefits from the formation of an FTA between the other two countries:

$$\Delta w_{l'}\left(\widehat{sl} - \Phi\right) > 0 \text{ for all } \theta \tag{20}$$

Second, as under the WTO-consistent benchmark case, the smaller country benefits from the formation of an FTA between the two larger countries only when the degree of asymmetry is

sufficiently small:

$$\Delta w_s\left(\widehat{ll'} - \Phi\right) > 0 \text{ when } \theta < \theta_s(\widehat{ll'} - \Phi) \quad (21)$$

A comparison of conditions in (14) and (21) yields

$$\theta_s(\widehat{ll'} - \Phi) < \theta_s(ll' - \Phi)$$

implying that the formation of $\langle\widehat{ll'}\rangle$ is *less likely* to be Pareto improving relative to $\langle ll'\rangle$. We can establish the proposition below:

Proposition 3: *(i) Relative to no agreement* $\langle\Phi\rangle$*, an unconstrained FTA between two asymmetric countries* $\langle\widehat{sl}\rangle$ *is necessarily Pareto-improving whereas the unconstrained FTA between the two larger importers* $\langle\widehat{ll'}\rangle$ *is Pareto-improving only when* $\theta < \theta_s(\widehat{ll'} - \Phi)$.

(ii) The freedom to set positive internal tariffs in a coordinated fashion makes the FTA between two asymmetric partners more likely to be Pareto-improving while the opposite is true for the FTA between the two larger importers.

We are now ready to derive equilibria under the game of unconstrained preferential liberalization. First note, it is immediate from part (*i*) of Lemma 3 that any two countries have an incentive to jointly deviate from their respective announcements under $\langle\Phi\rangle$ to announcement profiles leading to a bilateral FTA. Since this deviation is self-enforcing, $\langle\Phi\rangle$ is not a CPNE.

Next, consider the announcement profiles leading to $\langle\widehat{ll'}\rangle$. It is immediate from part (*iii*) and part (*v*) of the Lemma 3 that, taking the announcement profile of a large country (say l') as given, country s and either of the large member countries (say l) have incentives to jointly deviate from their respective announcements $\{\phi,\phi\}$ and $\{\phi,l'\}$ to $\{l,\phi\}$ and $\{s,l'\}$, leading to a deviation from $\langle\widehat{ll'}\rangle$ to $\langle\widehat{lh}\rangle$ and this deviation is self enforcing. As a result, the announcement profile leading to $\langle\widehat{ll'}\rangle$ is never a CPNE.

Consider now the announcement profile leading to global free trade $\langle F\rangle$. As in the benchmark case, note from part (*iii*) of Lemma 2 that any deviation (unilateral or coalitional) from $\langle F\rangle$ does not involve country s. Thus, if there exists a coalitional deviation, it must be by countries l and l'. Similar to the benchmark WTO case, when countries have the ability to set positive internal tariffs, large countries still have the incentive to exclude the small country. In other words, taking country s' announcement as fixed at $\{l,l'\}$, countries l and l' have incentives to jointly deviate from their respective announcements $\{s,l'\}$ and $\{s,l\}$ to $\{\phi,l'\}$ and $\{\phi,l\}$ in order exclude country s from a free trade network when country s is sufficiently small:

$$\Delta w_l\left(F - \widehat{ll'}\right) < 0 \text{ when } \theta > \theta_l(F - \widehat{ll'}) \quad (22)$$

The following result is based on the comparison of the exclusion incentives contained in (16) and (22):

Lemma 4: *The larger importers have a stronger incentive to exclude the smaller country from their mutual trade agreement under the unconstrained preferential liberalization scenario where they impose positive internal tariffs on each other relative to the WTO-consistent benchmark where they are required to fully liberalize internal trade:* $\theta_l(F - \widehat{ll'}) < \theta_l(F - ll')$.

We next argue that, as under the benchmark WTO case, the flexible nature of FTAs ensures that the exclusion incentive goes unexercised even when countries are able to impose positive internal tariffs on each other. To see why, suppose each country announces in favor of an FTA with both its trading partners. Part (*v*) of Lemma 2 informs us that a hub and spoke regime $\langle\widehat{lh}\rangle$ is a preferred regime for the hub country relative to being a member under $\langle\widehat{ll'}\rangle$. It follows then that, holding constant the announcement of the excluded small country at $\sigma_s = \{l, l'\}$, each member of the deviating coalition (l or l') has an incentive *to alter its announcement to form a separate FTA with the excluded country*. As a result, the original coalitional deviation of countries l and l' is not self-enforcing and thus the lure of a hub and spoke trading arrangement ends up undermining the exclusion incentives as before.

Next, taking country s' announcement as fixed $\{l, l'\}$, countries l and l' have no incentives to jointly deviate from their respective announcements of $\{s, l'\}$ and $\{s, l\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$, leading to a deviation from $\langle F\rangle$ to $\langle\Phi\rangle$. As before, the only possible self-enforcing deviation is the unilateral deviation of the either large importer from free trade. To this end, we find that, when the degree of asymmetry is sufficiently large, a large country (say l) has an incentive to unilaterally deviate from its announcement $\{s, l'\}$ to an announcement leading to a hub and spoke regime where the small country or the other large country is a hub and it itself is a spoke:

$$\Delta w_l\left(F - \widehat{sh}\right) < 0 \text{ when } \theta > \theta_l(F - \widehat{sh}) \tag{23}$$

and

$$\Delta w_l\left(F - \widehat{l'h}\right) < 0 \text{ when } \theta > \theta_l(F - \widehat{l'h}) \tag{24}$$

where $\theta_l(F - sh) < \theta_l(F - l'h)$. Then, the unilateral deviation incentive that remains to be examined is the unilateral deviation of a large country (say l) from $\{s, l'\}$ to $\{\phi, \phi\}$:

$$\Delta w_l\left(F - \widehat{sl'}\right) = \Delta w_{l'}\left(F - \widehat{sl}\right) < 0 \text{ when } \theta > \theta_l(F - \widehat{sl'}) \tag{25}$$

We find that $\theta_l(F - \widehat{sl'}) < \theta_l(F - \widehat{sh})$ holds and thus the announcement profile leading to $\langle F\rangle$ is CPNE when $\theta \leq \theta_l(F - \widehat{sl'})$.

We next examine the hub and spoke regimes. From part (*iv*) of Lemma 3, we know that a large spoke country (say l) under $\langle\widehat{sh}\rangle$ and $\langle\widehat{l'h}\rangle$ has an incentive to unilaterally deviate from its respective announcements $\{s, \phi\}$ and $\{\phi, l'\}$ to $\{\phi, \phi\}$ and $\{\phi, \phi\}$, leading to a deviation from $\langle\widehat{sh}\rangle$ to $\langle\widehat{sl'}\rangle$ and from $\langle\widehat{l'h}\rangle$ to $\langle\widehat{sl'}\rangle$ when the smaller country is sufficiently small and $\theta_l(\widehat{l'h} - \widehat{sl'}) < \theta_l(\widehat{sh} - \widehat{sl'})$. Moreover, when $\theta < \theta_l(\widehat{l'h} - \widehat{sl'})$, the joint announcement deviations of small and large countries leading to deviations from hub and spoke regimes to free trade are self-enforcing. Thus, the announcement profiles leading to any hub and spoke regime is never a CPNE.

The only remaining candidate for CPNE is the announcement profile leading to $\langle\widehat{sl}\rangle$. We know from part (*i*) of Lemma 3 that no country has an incentive to unilaterally deviate from its announcement leading to a deviation from $\langle\widehat{sl}\rangle$ to $\langle\Phi\rangle$. Second, we know from part (*v*) that the coalitional announcement deviation leading to a deviation from $\langle\widehat{sl}\rangle$ and $\langle\widehat{ll'}\rangle$ is not self-enforcing since the common member country (l here) always has an incentive to further deviate to become the hub country, taking the announcement of its complement fixed. Third, note from the above discussion that the coalitional announcement deviation leading a deviation from $\langle\widehat{sl}\rangle$ to $\langle F\rangle$ is self-enforcing only when $\theta \leq \theta_l(F - \widehat{sl'})$. When $\theta > \theta_l(F -$

$\widehat{sl'}$) holds, the non-member country l' has no incentive to engage in any coalitional announcement deviations that lead to a deviation from $\langle\widehat{sl}\rangle$ to $\langle\widehat{sh}\rangle$ or from $\langle\widehat{sl}\rangle$ to $\langle\widehat{lh}\rangle$. As a result, we argue that the announcement profile leading to $\langle\widehat{sl}\rangle$ is a CPNE when $\theta \geq \theta_l(F - \widehat{sl'})$.

The following proposition can now be stated:

Proposition 4: *The equilibria of the game of unconstrained preferential liberalization wherein FTA member countries coordinate their internal tariffs before setting their individually optimal external tariffs are as follows: if* $\theta \leq \theta_l(F - \widehat{sl'})$, *global free trade is the equilibrium outcome; otherwise, the asymmetric FTA* $\langle\widehat{sl}\rangle$ *(or* $\langle\widehat{sl'}\rangle$*) is the equilibrium outcome.*

A comparison of Propositions 2 and 4 yields the following result:

Proposition 5: *(i) For* $\theta \leq \theta_l(F - sl')$, *the equilibrium outcome is global free trade whether or not FTA members are required to practice free internal trade; whereas for* $\theta_l(F - sl') < \theta \leq \theta_l(F - \widehat{sl'})$, *it is the equilibrium only if FTA members are free to set positive internal tariffs on each other.*

(ii) When global free trade is out of reach, i.e. when $\theta > \theta_l(F - \widehat{sl'})$, *the free internal trade requirement of the WTO increases world welfare by yielding (weakly) lower global tariffs.*

The above proposition argues that, when the degree of endowment asymmetry is sufficiently small, global free trade arises regardless of whether or not FTA members are required to engage in free internal trade. However, when the degree of endowment asymmetry is moderate, global free trade arises *only when* FTA members are free to impose positive internal tariffs on each other. In other words, *the free internal trade requirement of GATT's Article XXIV hinders the cause of global free trade*. To understand this result, we should first note that the viability of global free trade is determined by the unilateral deviation incentive of one of the larger importers regardless of whether FTA members are required to engage in free internal trade or not. Due to the presence of tariff complementarity, the freedom to set positive internal tariffs leads FTA members to impose higher external tariffs which in turn makes it less attractive for one of the larger importers to opt out of global free trade – i.e. its incentive to free ride on the external trade liberalization of FTA members without having to offer any trade liberalization of its own is reduced. Finally, when global free trade is out of reach, the free internal trade requirement of Article XXIV acts as a disciplining device in a tariff-ridden world and it helps protect the interest of non-member country by leading FTA members to adopt lower external tariffs. Thus, our overall message is as follows: when circumstances are such that achieving complete global free trade is not possible, the free internal trade requirement of Article XXIV increases world welfare by reducing both internal

and external tariffs of FTAs but, at the same time, it also reduces the likelihood of reaching global free trade. From a practical perspective, given the multitude of ways in which countries can prevent the obtainment of global free trade, it would seem that the beneficial effects of Article XXIV's free internal trade requirement for FTAs are likely to be of greater real-world relevance than their negative effect on the prospects of achieving global free trade.

Ornelas (2005a) uses an oligopoly model of trade with political economy considerations to study related issues to our result here. While trade agreements are not endogenously determined in Ornelas (2005a), tariff complementarity and free riding incentives also play an important role in his analysis. In his model, the decline in external tariffs of members following FTA formation benefits the non-member country and when the degree of size asymmetry is sufficiently large, it can induce the large importing country to withdraw its support from multilateral trade agreement, such as global free trade.

5 Further analysis

In what follows, we extend our analysis in three important directions. First, we consider a scenario where FTA members have to abide by a ceiling on their internal tariffs as opposed to having to eliminate them completely. Second, we examine the consequences of allowing for tariff cooperation in the setting of MFN tariffs. Third, we allow all three countries to be asymmetric (as opposed to requiring two of them to be symmetric with respect to each other). While these extensions provide some interesting new insights, we find that our main results regarding the effects of the free internal trade requirement continue to hold.

5.1 Enforceability of Article XXIV

Under the unconstrained preferential liberalization scenario, we assume that FTA member countries jointly choose their internal tariffs before selecting their individually optimal external tariffs. We showed earlier that, due to the presence of tariff complementarity in our model, member countries set positive internal tariffs on each other. This has a subtle implication: if FTA members could set internal tariffs without restrictions, in a world of many countries and many goods, countries could form an FTA with every other country and set country-specific tariffs on every good. This would essentially imply the end of the MFN rule. Thus, it is useful to consider the role of the free internal trade requirement of Article XXIV and the extent of its enforceability by considering a scenario where Article XXIV establishes a *ceiling* on the internal tariffs of FTAs as opposed to calling for their outright elimination.

From our previous analysis, the following can be established: (*i*) when the ceiling on internal tariffs of an FTA is set to zero, we are in the WTO-consistent scenario, and (*ii*) when this ceiling is set above the optimal internal tariff, it becomes redundant and we are in the unconstrained preferential liberalization scenario. As a result, the ceiling only binds if it falls between zero and the optimal internal tariffs of FTA members. Over this binding range, we find that the free riding incentive – that is pivotal for the stability of global free trade – becomes a function of the institutionally given tariff ceiling.

Let $\bar{\tau}$ denote the ceiling facing the FTA's internal tariff where $\bar{\tau} \leq \min(\tau_{ij}^*,\ \tau_{ji}^*)$ so

that it binds for FTA members. The following result, represented in Figure 4, shows that our main result is robust to the existence of such a tariff ceiling:

Proposition 6: *Suppose the internal tariffs of an FTA are subject to a tariff ceiling* $\bar{\tau} \leq \min(\tau_{ij}^*, \tau_{ji}^*)$. *Then, global free trade is the equilibrium outcome whenever* $\theta \leq \theta_l(F - \widehat{sl'})$. *A larger country's incentive to remain in its FTAs decreases as the tariff ceiling becomes more binding (i.e. lower), i.e.* $\frac{\partial \theta_l(F-\widehat{sl'})}{\partial \bar{\tau}} > 0$.

The above proposition states that, even when an internal tariff ceiling exists and binds, the incentive of a larger importing country to free ride on the external trade liberalization of FTA members is pivotal to the stability of global free trade. In fact, this free riding incentive is a continuously increasing function of the internal tariff ceiling. As a result, *a lower FTA internal tariff ceiling makes it harder to achieve global free trade*. The intuition behind our previous results follows through: due to the presence of tariff complementarity, the ability to set positive internal tariffs (constrained by the internal tariff ceiling) leads FTA members to impose higher external tariffs relative to the case of free internal trade. This in turn makes it less attractive for one of the larger importers to opt out of global free trade.

5.2 Cooperation in MFN tariffs

Thus far, consistent with the widespread assumption in the PTA literature, we have assumed that countries set their MFN tariffs non-cooperatively. This poses a limitation given the fact that, under the GATT/WTO, countries not only form FTAs with each other, but also cooperate to some degree while setting their MFN tariffs. We now demonstrate that our main results continue to hold even when we allow for some degree of cooperation between countries during the setting of MFN tariffs.

Let μ denote the weight each country assigns to the welfare of other countries in setting its MFN tariff. Let

$$t_i^{\mu} \equiv \arg\max w_i(t_{ij}, t_{ik}) + \mu w_j + w_k] \text{such that } t_{ij} = t_{ik} \tag{26}$$

where

$$t_i^{\mu} = \frac{1}{2}\frac{(e_j + e_k)(1-\mu)}{(4-\mu)} \leq t_i^M = \frac{e_j + e_k}{8}$$

The case where countries set tariffs completely non-cooperatively arises when $\mu = 0$ while $\mu = 1$ captures full tariff cooperation. The latter case of complete cooperation is uninteresting because when $\mu = 1$, countries fully internalize the effects of their tariffs on their trade partners, and the optimal MFN tariff of each country ends up being equal to zero (which in turn eliminates any reason to form trade agreements). When $\mu \in [0,1/3]$, the

(partially) cooperative tariff t_i^{μ} lies between the optimal non-cooperative MFN tariff $t_i^M = t_i^{\mu}\big|_{\mu=0}$ and the optimal FTA external tariff that country i imposes on country k as a non-member (i.e. $t_{ik}^*(\tau_{ij}^*)$). Hereafter, we assume that $\mu \in [0,1/3]$, with the parameter μ capturing the degree of cooperation between countries.[18] We set this constraint on μ since one of our major insights is that lower internal tariffs of an FTA benefit the non-member by inducing FTA members to *reduce* their external tariffs. So in order for the FTA members to be able to react optimally to their internal tariffs, the bound MFN tariff rate cannot be too low. It is worth noting that the tariff t_i^{μ} can also be interpreted as the multilaterally negotiated tariff binding that countries have committed to previously, capping the maximum applied MFN tariff.[19]

As μ rises, the cooperative MFN tariffs decline, which limits the non-member country's ability to set its optimal MFN tariff while there is no limit on the optimal FTA external tariffs that FTAs members impose on it. As a result, the free riding incentive decreases under both scenarios, making global free trade more likely to emerge as an equilibrium outcome. To facilitate the statement of the formal result, let $\Delta w_i(r^c - v^c)$ denote the difference between country i's welfare under trade agreements r^c and v^c with cooperative MFN tariffs and free internal trade requirement: $\Delta w_i(r^c - v^c) \equiv w_i(r^c) - w_i(v^c)$. Similarly, $\Delta w_i(\hat{r}^c - \hat{v}^c)$ denotes the difference between country i's welfare under trade agreements $\hat{r}^c$ and $\hat{v}^c$ with cooperative MFN tariffs under the unconstrained preferential liberalization scenario: $\Delta w_i(\hat{r}^c - \hat{v}^c) \equiv w_i(\hat{r}^c) - w_i(\hat{v}^c)$. The critical threshold asymmetries for both $\theta_i(r^c - v^c)$ and $\theta_i(\hat{r}^c - \hat{v}^c)$ are then determined accordingly. The following proposition summarizes our findings:

Proposition 7: *Suppose countries partially cooperate in setting their MFN tariffs, where $\mu \in [0,1/3]$ denotes the weight that each country puts on the welfare of its trading partners (so that $t_i^C \leq t_i^M$). Then, the following holds:*

(i) When $\theta \leq \theta_l(F - sl'^c)$ global free trade is the equilibrium outcome regardless of whether FTA members are required to practice free internal trade.

(ii) When $\theta_l(F - sl'^c) < \theta \leq \theta_l(F - \widehat{sl'^c})$ global free trade is the equilibrium outcome only when FTA members are free to impose positive internal tariffs on each other.

(iii) As the weight that each country places on the welfare of its trading partners rises, the free riding incentive falls in both scenarios and the likelihood of obtaining global free trade increases: $\frac{\partial \theta_l(F - sl'^c)}{\partial \mu} > 0$ and $\frac{\partial \theta_l(F - \widehat{sl'^c})}{\partial \mu} > 0$.

The above proposition shows that our results are robust to relaxing the assumption that

[18] In an important recent paper, Olarreaga et al. (2018) show that more than three-quarters of WTO members' tariffs are set non-cooperatively.

[19] While the free internal trade requirement does not play any role in their models, recent working papers by Nken and Yildiz (2018) and Lake et al. (2018) investigate the implications of multilateral trade liberalization (i.e. continual reduction in tariff bindings) on static and dynamic incentives for PTA formation.

MFN external tariffs are set non-cooperatively. Under this initial assumption, the larger importing country as a non-member faces discrimination when exporting to the members' markets while benefiting from the tariff complementarity practiced by FTA members and from imposing its optimal tariff on both FTA members. Indeed, when the degree of country asymmetry is sufficiently large, the larger importing country has an incentive to unilaterally deviate from global free trade and stay outside of a bilateral FTA. When external MFN tariffs are set cooperatively, we find that the cooperative tariff is lower, $t_i^C \leq t_i^M$. This means that the non-member country, under a bilateral FTA, loses its ability to set its optimal non-cooperative tariff and is required to impose the cooperative tariff (i.e. tariff ceiling). On the other hand, the FTA members enjoy free access to each others' markets and are free to impose their optimal external tariffs on the non-member (when $\mu \in [0,1/3]$). This makes the discrimination faced by the FTA non-member more prominent thereby *weakening* its free riding incentive. As indicated by part (*iii*) of Proposition 7, regardless of the existence of the free internal trade requirement, a higher degree of cooperation in setting MFN external tariffs expands the range of endowment asymmetry over which global free trade is an equilibrium. The second part of Proposition 7 shows that our main result stays unchanged regardless of whether MFN tariffs are set cooperatively or non-cooperatively: the free internal trade requirement of Article XXIV makes it *less* attractive for the larger importing country to enter into trade agreements with the other two countries and thus reduces the likelihood of reaching global free trade since $\theta_l(F - sl'^C) < \theta \leq \theta_l(F - \widehat{sl'^C})$.

5.3 Greater degree of endowment asymmetry

Our core model considers an endowment structure where one country has a larger endowment of non-numeraire good than the other two countries. In this section, we show that our main results are robust to relaxing this current endowment pattern to allow for all three countries to be asymmetric. Specifically, let there be a medium importing country in addition to the larger and smaller importing countries. We denote the larger importer country as l, the medium size importer as m and the smaller importer as s: $e_s = \theta_s e \geq e_m = \theta_m e \geq e_l = \theta_l e$ where θ_l is normalized to 1. Since all countries have asymmetric endowments, country l faces the largest import volume of protected goods under free trade (it imports $(e_s + e_m)/3$ units of good L) whereas country s faces the lowest import volume of such goods (it imports $(e_m + e_l)/3$ units of good S). As before, in order to guarantee non-negative exports and positive tariffs under all regimes in all scenarios, we assume that $\frac{5}{4} \geq \theta_s \geq \theta_m \geq 1$ holds hereafter. Let $\theta_i(r - v)$ denote the larger country's critical endowment threshold, as a function of the medium country's endowment, at which country i is indifferent between regimes r and v.

We can show the following:

Proposition 8: *Suppose there are three asymmetric countries:* $\frac{5}{4} \geq \theta_s \geq \theta_m \geq 1$. *Then, the following holds:*

(i) When $\theta_s \leq \theta_l(F - sm)$ *global free trade is the equilibrium outcome regardless of whether FTA members are required to practice free internal trade or not.*

(ii) When $\theta_l(F - sm) < \theta_s \leq \theta_l(F - \widehat{sm})$ *global free trade is the equilibrium outcome only when FTA members are free to impose positive internal tariffs on each other.*

(iii) If global free trade is out of reach, the free internal trade requirement improves welfare.

The first part of Proposition 8 states that, when the degree of endowment asymmetry is sufficiently small, global free trade is a stable outcome regardless of whether free internal trade is required or not. However, when the degree of endowment asymmetry exceeds a certain threshold, global free trade arises only when the free internal trade requirement does not bind. As before, with or without these requirements, the larger importing country's unilateral deviation incentive (free riding incentive) is critical for the stability of global free trade. Thus, the ability of smaller and medium importing countries to coordinate their internal tariffs before setting external tariffs under an FTA leads to smaller degree of tariff complementarity which reduces the larger country's incentive to unilaterally deviate and free ride on the trade liberalization of the member countries.

Finally, when global free trade is out of reach, the free internal trade requirement acts as a disciplining device for not only internal tariffs but also external tariffs due to tariff complementarity. Since the free internal trade requirement leads to deeper trade liberalization, its adoption leads to higher world welfare in a tariff-ridden world.

6 Conclusion

The core rule governing the formation of FTAs in the WTO is Article XXIV of the General Agreement of Tariffs and Trade (GATT). Under Article XXIV, countries entering into an FTA are required to: (a) eliminate trade restrictions on *substantially all trade* between themselves and (b) refrain from raising trade restrictions on non-member countries. In our competing exporters model, due to the existence of tariff complementarity, the second requirement of Article XXIV turns out to be non-binding and the fate of the outside countries ends up depending solely upon whether or not FTA members have to abide by the first condition, i.e., fully liberalize their internal trade.

To draw out the implications of requiring FTA members to eliminate tariffs on one another, we derive and contrast optimal tariffs and equilibrium trade agreements under two scenarios: under the WTO-consistent scenario, members are required to engage in free internal trade whereas under unrestricted preferential liberalization scenario members are free to impose non-zero internal tariffs on one another. Under both scenarios, the non-member is required to follow MFN. A comparison of these scenarios delivers several new insights. First, we show that the PTA members' incentive to maintain positive *internal* tariffs on each other depends on how they set their *external* tariffs. If PTA members set external tariffs independently, as they do in an FTA, they benefit from *not* eliminating their internal tariffs since doing so commits them to higher external tariffs. On the other hand, when external tariffs are coordinated – as they are under a CU – PTA members find it optimal to eliminate internal tariffs

so that the restriction on internal tariffs imposed by Article XXIV becomes moot.

Our second major result is rather surprising: requiring FTA members to eliminate internal tariffs *benefits* the non-member since it leads to *lower* external tariffs on the part of FTA members. In other words, it is the Article XXIV requirement of free internal trade amongst FTAs that ends up protecting the interest of the non-member as opposed to the Article's restriction on external tariffs imposed on FTA members. Indeed, we show that the free internal trade requirement can make it more likely that an FTA between two countries is Pareto-improving relative to a scenario where no trade agreements exist.

Since our analysis derives equilibrium agreements in a game in which all countries are free to form trade agreements with one another, we are able to speak to the consequences of the free internal trade requirement of Article XXIV for the likelihood of achieving global free trade. Our major finding is that this requirement makes it *harder* to achieve global free trade by limiting the negative impact of an FTA on the non-member country: due to tariff complementarity, lower internal tariffs within an FTA also imply lower external tariffs. By not entering into a trade agreement with FTA members, the non-member country remains free to impose its optimal import tariffs on them while itself facing relatively lower tariffs in their markets. Thus, it is possible that the free internal trade requirement of Article XXIV *facilitates some degree of free-riding in the WTO system* by making it possible for non-member countries to benefit from reductions in external tariffs of FTA members without having to reciprocate with tariff cuts of their own. However, while the free internal trade requirement of Article XXIV reduces the likelihood of obtaining global free trade, it also increases welfare by lowering tariffs world-wide when global free trade is simply out of reach.

Finally, while we have examined the implications of the free internal trade requirement facing PTAs for both FTAs and CUs, our approach has abstracted from the endogenous choice between these two types of PTAs. This is an important question for future research.

7 Appendix

In this Appendix we provide all supporting calculations and proofs.

7.1 Supporting calculations

We begin by reporting welfare levels as functions of an arbitrary tariff vector. Then, we report the optimal tariffs under each trade regime. Using the welfare and tariff levels reported below, we can easily obtain the formulae for optimum welfare levels under all possible regimes. Lemmas 1, 2, 3, 4, 5 and the various inequalities reported in the main text follow from a direct application of the relevant formulae.

7.1.1 Welfare levels

We report welfare levels for country i under a trade regime r as a function of an arbitrary tariff vector $\mathbf{t}(r)$ where $\mathbf{t}(r) = (t_{ij}(r), t_{ik}(r))$:

$$w_i(r) = \sum_z CS_i^z(r) + \sum_z PS_i^z(r) + TR_i(r)$$

where

$$\sum_z CS_i^z(r) = \frac{1}{2}\left[\left(\frac{e_j + e_k - t_{ij}(r) - t_{ik}(r)}{3}\right)^2 + \left(\frac{e_i + e_k + 2t_{ji}(r) - t_{jk}(r)}{3}\right)^2 + \left(\frac{e_i + e_j + 2t_{ki}(r) - t_{kj}(r)}{3}\right)^2\right]$$

$$\sum_z PS_i^z(r) = \frac{e_i[6\alpha - 2e_i - e_j - e_k + t_{jk}(r) + t_{kj}(r) - 2t_{ji}(r) - 2t_{ki}(r)]}{3}$$

and

$$TR_i(r) = \frac{t_{ij}(r)[2e_j - e_k + t_{ik}(r) - 2t_{ij}(r)]}{3} + \frac{t_{ik}(r)[2e_k - e_j + t_{ij}(r) - 2t_{ik}(r)]}{3}.$$

7.1.2 Optimal Tariffs

Next, we report the optimal tariffs under each regime and provide supporting calculations for our tariff discussion in the text. Country i's optimal MFN tariff is

$$t_i^{\phi} \equiv Arg\max w_i(\Phi) = \frac{e_j + e_k}{8} \quad (27)$$

Next, we examine the FTA member tariffs. First we show that, holding everything else constant, the non-member country loses as internal tariffs of an FTA decline:

$$\frac{\partial w_k(ij)}{\partial \tau_{ij}} = \frac{2(e_k - t_{ik}) - (e_j - \tau_{ij})}{9} > 0$$

Suppose now that external tariffs are optimally chosen. Then, we find the following optimal external tariff as a function of internal tariff between member countries:

$$t_{ik}(ij) = \frac{5e_k - 4e_j + 7\tau_{ij}}{11}$$

Note that the tariff complementarity holds:

$$\frac{\partial t_{ik}(ij)}{\partial \tau_{ij}} = \frac{7}{11} > 0$$

We next show that, when external tariffs are optimally chosen by FTA members, we obtain:

$$\frac{\partial w_k(ij)}{\partial \tau_{ij}} = -\frac{4e_k - \tau_{ij} - e_j}{121} < 0$$

If countries could coordinate internal tariffs before setting their individually optimum external tariffs, FTA members can partially internalize the effects of their tariffs on one another: $(\tau_{ij}, \tau_{ji}) \equiv \arg\max \left[w_i(ij) + w_j(ij)\right]$:

$$\tau_{ij} = \frac{3e_j - e_k}{63} > 0$$

Then the optimal external tariff is as follows:

$$t_{ik}(ij) = \frac{4e_k - 3e_j}{9} \quad (28)$$

Under free internal trade, the optimum external tariff under an FTA (and the optimal spoke's tariff under a hub and spoke regime) is immediate:

$$t_{ik}(ij) = t_{ik}(jh) = \frac{5\ e_k - 4e_j}{11}$$

Under a CU, we found the following optimum external tariffs as a function of the internal tariffs:

$$t_{ik}(ij^u) = \frac{2\ e_k - e_j}{5} + \frac{\tau_{ij}}{2} \quad (29)$$

Note that, while it is weaker relative to an FTA game, the tariff complementarity still holds:

$$\frac{\partial t_{ik}(ij^u)}{\partial \tau_{ij}} = \frac{1}{2} > 0$$

We find that it is optimum for CU members to eliminate internal tariffs:

$$\frac{\partial [w_i(ij^u) + w_j(ij^u)]}{\partial \tau_{ij}} = -\frac{\tau_{ij}}{2} < 0$$

As a result, the following jointly optimal external tariffs under $\langle ij^u \rangle$ obtain:

$$t_{ik}(ij^u) = \frac{2\ e_k - e_j}{5} \quad (30)$$

Note that we obtain higher external tariffs under a CU relative to an FTA: $t_{ik}(ij^u) > t_{ik}(ij)$.

7.2 Proofs of Lemmas and Propositions

Note that the proof of Lemma 1 is immediate from the optimal tariff discussion above.

Proof of Lemma 2

Using the above welfare formulae (as functions of an arbitrary tariff vector) and plugging the above optimum tariffs into them, it is straightforward to show the following inequalities:

Part (i): $\Delta w_l(ll' - sl) > 0$ for all $1 \leq \theta \leq 5/4$.

Part (ii): $\Delta w_s(sl - \Phi) > 0$, $\Delta w_s(sh - sl) > 0$, and $\Delta w_s(F - lh) > 0$ hold for all $1 \leq \theta \leq 5/4$ while $\Delta w_s(lh - ll') > 0$ only when $\theta > \theta_s(lh - ll') \cong 1.03$.

Part (iii): $\Delta w_{l'}(lh - sl) < 0$ and $\Delta w_l(l'h - sl') < 0$ for all $1 \leq \theta \leq 5/4$ while $\Delta w_s(lh - ll') > 0$ only when $\theta > \theta_s(lh - ll') \cong 1.03$.

Part (iv): $\Delta w_i(ih-\Phi)>0$, $\Delta w_i(ih-F)>0$ and $\Delta w_i(ih-ij)>0$ for all for all $1\leq\theta\leq 5/4$ and $i=s,l,l'$.

Proof of Proposition 1

Using the above welfare formulae (as functions of an arbitrary tariff vector) and plugging the above optimum tariffs into them, it is straightforward to show that $\Delta w_s(sl-\Phi)>0$ holds for all $1\leq\theta\leq 5/4$ while $\Delta w_l(sl-\Phi)>0$ only when $\theta<\theta_l(sl-\Phi)\cong 1.24$. Similarly, we obtain $\Delta w_l(ll'-\Phi)>0$ for all $1\leq\theta\leq 5/4$ while $\Delta w_s(ll'-\Phi)>0$ only when $\theta<\theta_s(ll'-\Phi)\cong 1.09$.

Proof of Proposition 2

Using the results from Lemma 2, the discussion in the main text and the following inequalities, it is straightforward to prove Proposition 2:

- $\Delta w_l(F-ll')<0$ when $\theta>\theta_l(F-ll')\cong 1.085$;
- $\Delta w_l(F-l'h)=\Delta w_{l'}(F-lh)<0$ when $\theta>\theta_l(F-l'h)\cong 1.18$;
- $\Delta w_l(F-sl')=\Delta w_{l'}(F-sl)<0$ when $\theta>\theta_l(F-sl')\cong 1.081$;
- $\Delta w_s(lh-ll')>0$ when $\theta>\theta_s(lh-ll')\cong 1.03$.

Proof of Lemma 3

Using the above welfare formulae (as functions of an arbitrary tariff vector) and plugging the above optimum tariffs (without free internal trade requirement) into them, it is straightforward to show the following inequalities:

Part (i): $\Delta w_s(\widehat{sl}-\Phi)>0$, $\Delta w_l(\widehat{sl}-\Phi)>0$, $\Delta w_l(\widehat{ll'}-\Phi)>0$ for all $1\leq\theta\leq 5/4$.

Part (ii): $\Delta w_l(\widehat{ll'}-\widehat{sl})>0$ for all $1\leq\theta\leq 5/4$.

Part (iii): $\Delta w_s(\widehat{sl}-\Phi)>0$, $\Delta w_s(\widehat{sh}-\widehat{sl})>0$, $\Delta w_s(F-\widehat{lh})>0$ and $\Delta w_s(lh-ll')>0$ hold for all $1\leq\theta\leq 5/4$.

Part (iv): $\Delta w_l(\widehat{l'h}-\widehat{sl'})<0$ when $\theta>\theta_l(\widehat{l'h}-\widehat{sl'})\cong 1.029$ and $\Delta w_l(\widehat{sh}-\widehat{sl'})<0$ when $\theta>\theta_l(\widehat{l'h}-\widehat{sl'})\cong 1.037$.

Part (v): $\Delta w_i(\widehat{ih}-\Phi)>0$ and $\Delta w_i(\widehat{ih}-\widehat{ij})>0$ for all for all $1\leq\theta\leq 5/4$ and $i=s,l,l'$.

Proof of Proposition 3

Along with the first part of Lemma 3, using the above welfare formulae (as functions of an arbitrary tariff vector) and plugging the above optimum tariffs (without free internal trade requirement) into them, it is straightforward to show that $\Delta w_{l'}(\widehat{sl}-\Phi)>0$ holds for all $1\leq\theta\leq 5/4$ while $\Delta w_s(\widehat{ll'}-\Phi)>0$ only when $\theta<\theta_s(\widehat{ll'}-\Phi)\cong 1.076$.

Proof of Lemma 4

Using the above welfare formulae (as functions of an arbitrary tariff vector) and

plugging the above optimum tariffs (with and without free internal trade requirement) into them, it is straightforward to show that $\theta_l(F-\widehat{ll'}) \cong 1.082 < \theta_l(F-ll') \cong 1.085$.

Proof of Proposition 4

Using the results from Lemmas 3 and 4, the discussion in the main text and the following inequalities, it is straightforward to prove Proposition 4:

- $\Delta w_l(F-\widehat{ll'}) < 0$ when $\theta > \theta_l(F-\widehat{ll'}) \cong 1.082$.
- $\Delta w_l(F-\widehat{l'h}) = \Delta w_{l'}(F-\widehat{lh}) < 0$ when $\theta > \theta_l(F-\widehat{l'h}) \cong 1.138$.
- $\Delta w_l(F-\widehat{sh}) = \Delta w_l(F-\widehat{sh}) < 0$ when $\theta > \theta_l(F-\widehat{sh}) \cong 1.130$.
- $\Delta w_l(F-\widehat{sl'}) = \Delta w_{l'}(F-\widehat{sl}) < 0$ when $\theta > \theta_l(F-\widehat{sl'}) \cong 1.097$.

Proof of Proposition 5

The proof is immediate from the proofs of Propositions 2 and 4.

Proof of Proposition 6

Using the above welfare formulae (as functions of an arbitrary tariff vector) and plugging in the optimum external tariffs as functions of the exogenous ceiling on the FTA's internal tariff $\bar{\tau}$, it is straightforward to show that one of the larger importing countries (say l) has an incentive to unilaterally deviate from $\{s,l'\}$ to $\{\phi,\phi\}$, leading to a deviation from $\langle F\rangle$ to $\widehat{\langle sl'\rangle}$: $\Delta w_l(F-\widehat{sl'}) = \Delta w_{l'}(F-\widehat{sl}) < 0$ when $\theta > \theta_l(F-\widehat{sl'})$. Note that this deviation is self-enforcing and there exists no other self-enforcing deviation whenever $\theta \leq \theta_l(F-\widehat{sl'})$. As a result, the announcement profile leading to global free trade $\langle F\rangle$ is a CPNE when $\theta \leq \theta_l(F-\widehat{sl'})$ and $\theta_l(F-\widehat{sl'})$ is an increasing function of $\bar{\tau}$ as indicated in Figure 4.

Proof of Proposition 7

Denoting the extra weight each country assigns to the welfare of other countries in setting its MFN tariff by μ, we obtain the optimal cooperative MFN tariff as follows: $t_i^{\mu} = \frac{1}{2}\frac{(e_j+e_k)(1-\mu)}{(4-\mu)}$. As discussed in the text, we restrict our attention to the range $\mu \in [0,1/3]$. Under the WTO-consistent scenario and the unconstrained preferential liberalization scenario, it is straightforward to show that one of the large importing countries (say l) has an incentive to unilaterally deviate from $\{s,l'\}$ to $\{\phi,\phi\}$, leading to a deviation from $\langle F\rangle$ to an FTA between the other two countries: $\Delta w_l(F-sl'^c) < 0$ when $\theta > \theta_l(F-sl'^c)$ and $\Delta w_l(F-\widehat{sl'^c}) < 0$ when $\theta > \theta_l(F-\widehat{sl'^c})$. Note that under both scenarios these unilateral deviations are self-enforcing and there exists no other self-enforcing deviation when $\theta \leq \theta_l(F-sl'^c)$ and $\theta \leq \theta_l(F-\widehat{sl'^c})$, respectively. As a result, the announcement profile leading to global free trade $\langle F\rangle$ is a CPNE under the WTO-consistent scenario when $\theta \leq \theta_l(F-sl'^c)$ while it is a CPNE under the unconstrained preferential liberalization scenario when $\theta \leq \theta_l(F-\widehat{sl'^c})$. It is straightforward to show that the free internal trade requirement

of Article XXIV makes it *more* attractive for the larger importing country to free ride on trade liberalization by the other two countries and thus reduces the likelihood of reaching global free trade since $\theta_l(F - sl'^c) < \theta_l(F - \widehat{sl'^c})$. As indicated in Figure 5, both $\theta_l(F - sl'^c)$ and $\theta_l(F - \widehat{sl'^c})$ are increasing functions of the degree of cooperation between countries μ.

Proof of Proposition 8

Let $e_s = \theta_s e \geq e_m = \theta_m e \geq e_l = \theta_l e$ where θ_l is normalized to 1. Using the above welfare formulae (as functions of an arbitrary tariff vector) and plugging the above optimum tariffs into them, it is straightforward to show that, under both WTO consistent scenario and unconstrained preferential liberalization scenario, the larger importing country l has an incentive to unilaterally deviate from $\{s, m\}$ to $\{\phi, \phi\}$, leading to a deviation from $\langle F \rangle$ to an FTA between the other two countries: $\Delta w_l(F - sm) < 0$ when $\theta > \theta_l(F - sm)$ and $\Delta w_l(F - \widehat{sm}) < 0$ when $\theta > \theta_l(F - \widehat{sm})$. Note under both scenarios that these unilateral deviations are self-enforcing and there exists no other self-enforcing deviation when $\theta \leq \theta_l(F - sm)$ and $\theta \leq \theta_l(F - \widehat{sm})$, respectively. As a result, the announcement profile leading to global free trade $\langle F \rangle$ is a CPNE under the WTO-consistent scenario when $\theta \leq \theta_l(F - sm)$ while it is a CPNE under the unconstrained preferential liberalization scenario when $\theta \leq \theta_l(F - \widehat{sm})$ holds. Note also from Figure 6 that $\theta_l(F - \widehat{sm}) > \theta_l(F - sm)$. As argued before, when global free trade is out of reach, the free internal trade requirement leads to lower internal and external tariffs on the part of FTAs and therefore increases world welfare.

References

[1] Aghion, Philippe, Pol Antràs, and Elhanan Helpman 2007. " Negotiating free trade." *Journal of International Economics* 73: 1–30.

[2] Bagwell, K. and Robert. W. Staiger, 1998. " Will preferential agreements undermine the multilateral trading system?" *The Economic Journal*, 108(449), pp.1162-1182.

[3] Bagwell, Kyle and Robert. W. Staiger, 1999a. " Regionalism and multilateral tariff cooperation." In John Piggott and Allan Woodland, eds, *International Trade Policy and the Pacific Rim*, London: MacMillan.

[4] Bagwell, Kyle and Robert W. Staiger, 1999b. " An economic theory of GATT." *American Economic Review* 89(1): 215-48.

[5] Bagwell, Kyle, Chad P. Bown, and Robert W. Staiger, 2016. " Is the WTO passé?" *Journal of Economic Literature* 54(4): 1125–1231.

[6] Bernheim, Douglas B., Bezalel Peleg and Michael Whinston, 1987. " Coalition-proof Nash Equilibria I. Concepts." *Journal of Economic Theory* 42: 1-12.

[7] Bhagwati, Jagdish, 1991. *The World Trading System at Risk*. Princeton University Press, Princeton, NJ.

[8] Bohara, A.K., Gawande, K. and Sanguinetti, P., 2004. " Trade diversion and declining tariffs: evidence from Mercosur." *Journal of International economics*, 64(1), pp.65-88.

[9] Bond, E.W., Riezman, R.G. and Syropoulos, C., 2004. " A strategic and welfare theoretic analysis of free trade areas." *Journal of International Economics*, 64(1), pp.1-27.

[10] Cadot, Olivier, Jaime De Melo, and Marcelo Olarreaga. 1999. " Regional integration and lobbying for tariffs against nonmembers." *International Economic Review* 40(3): 635-658.

[11] Calvo-Pardo, H., Freund, C. and Ornelas, E., 2009. " The ASEAN free trade agreement: impact on trade flows and external trade barriers." The World Bank.

[12] Caplin, Andrew, and Kala Krishna, 1988. " Tariffs and the most favored nation clause: a game theoretic approach." *Seoul Journal of Economics* 1, 267-289.

[13] Crivelli, P., 2016. " Regionalism and falling external protection in high and low tariff members." *Journal of International Economics*, 102, pp.70-84.

[14] Damuri, Yose Rizal, 2012. " How preferential are preferential trade agreements? Analysis of product exclusions in PTAs." Graduate Institute Centre for Trade and Economic Integration Working Paper, 2012-3.

[15] Estevadeordal, A., Freund, C. and Ornelas, E., 2008. " Does regionalism affect trade liberalization toward nonmembers?." *The Quarterly Journal of Economics*, 123(4), pp.1531-1575.

[16] Freund, Caroline 2000. " Multilateralism and the endogenous formation of preferential trade agreements." *Journal of International Economics* 52(2): 359-376.

[17] Freund, Caroline L., and Emanuel Ornelas, 2010. " Regional trade agreements." *Annual Review of Economics* 2: 139-166.

[18] Furusawa,Taiji and Hideo Konishi, 2007. " Free trade networks." *Journal of International Economics* 72: 310–335.

[19] Goyal, Sanjeev and Sumit Joshi, 2006. " Bilateralism and free trade." *International Economic Review* 47: 749–778.

[20] Johnson, H.G., 1965. " An Economic Theory of protectionism, tariff bargaining, and the formation of customs unions." *The Journal of Political Economy* 72: 256-83.

[21] Kennan, John and Raymond Riezman, 1990. " Optimal tariff equilibria with customs unions." *Canadian Journal of Economics* 23(1): 70-83.

[22] Krishna, Pravin,1998. " Regionalism and multilateralism: a political economy approach." *Quarterly Journal of Economics* 113: 227–51.

[23] Lake, James, Moise Nken and Halis M. Yildiz, 2018." Tariff bindings and the dynamic formation of preferential trade agreements." Mimeo.

[24] Liu, Xuepeng. 2010. " Testing conflicting political economy theories: full-fledged versus partial-scope regional trade agreements." *Southern Economic Journal* 77(1): 78-103.

[25] Ludema, R.D. and Mayda, A.M., 2009. " Do countries free ride on MFN?" *Journal of International Economics* 77(2), pp.137-50.

[26] Ludema, R.D. and Mayda, A.M., 2013. " Do terms-of-trade effects matter for trade agreements? Theory and evidence from WTO countries." *The Quarterly Journal of Economics* 128(4): 1837-93.

[27] Maggi, Giovanni, 2014. " International trade agreements." In *Handbook of international Economics*. Vol. 4. Elsevier. 317-390.

[28] Mai, J. and Stoyanov, A., 2015. " The effect of the Canada-US free trade agreement on Canadian multilateral trade liberalization." *Canadian Journal of Economics*, 48(3), pp.1067-1098.

[29] Menon, Jayant, 2010. " Dealing with the proliferation of bilateral free trade agreements." *World Economy*, 32(10), pp.1381-1407.

[30] Missios, Paul, Kamal Saggi and Halis M. Yildiz, 2016. " External trade diversion, exclusion incentives and the nature of preferential trade agreements." *Journal of International Economics* 99: 105-19.

[31] Mrázová, M., Vines, D. and Zissimos, B., 2013. " Is the GATT/WTO's Article XXIV bad?" *Journal of International Economics*, 89(1), pp.216-232.

[32] Nken, Moise and Halis M. Yildiz, 2018." Implications of multilateral tariff bindings on the extent of preferential trade agreement formations." Mimeo.

[33] Olarreaga, Marcelo and Isidro Soloaga, 1998. " Endogenous tariff formation: the case of Mercosur." *World Bank Economic Review* 12(2): 297-320.

[34] Olarreaga, Marcelo, M., Nicita, Alessandro and Peri Silva, 2018. " Cooperation in WTO's tariff waters." *Journal of Political Economy* 126 (3): 1302-1338.

[35] Ornelas, Emanuel, 2005a. " Trade creating free trade areas and the undermining of multilateralism." *European Economic Review* 49: 1717-35.

[36] Ornelas, Emanuel, 2005b. " Rent destruction and the political viability of free trade agreements." *Quarterly Journal of Economics* 120: 1475-1506.

[37] Ornelas, Emanuel, 2007. " Exchanging market access at the outsiders' expense: the case of customs unions." *Canadian Journal of Economics* 40(1): 207-24.

[38] Ossa, Ralph, 2011. " A new trade theory of GATT/WTO negotiations." *Journal of Political Economy*, 119(1), 122-152.

[39] Richardson, Martin, 1993. " Endogenous protection and trade Diversion." *Journal of International Economics* 34(3-4): 309-24.

[40] Richardson, Martin, 1995. " Tariff revenue competition in a free trade area."

European Economic Review 39(7): 1429-37.

[41] Saggi, Kamal, 2004. " Tariffs and the most favored nation clause." *Journal of International Economics* 63(2), 341-68.

[42] Saggi, Kamal and Halis M. Yildiz, 2009. "Optimal tariffs of preferential trade agreements and the tariff complementarity effect." *Indian Growth and Development Review*, 2(1), pp.5-17.

[43] Saggi, Kamal and Halis M. Yildiz, 2010. " Bilateralism, multilateralism, and the quest for global free trade." *Journal of International Economics* 81: 26-37.

[44] Saggi, Kamal and Halis M. Yildiz, 2011. "Bilateral trade agreements and the feasibility of multilateral free trade." *Review of International Economics*, 19(2), pp.356-373.

[45] Saggi, Kamal, Alan Woodland and Halis M. Yildiz, 2013. " On the relationship between preferential and multilateral trade liberalization: the case of customs unions." *American Economic Journal: Microeconomics* 5(1): 63-99.

[46] Seidmann, Daniel, 2009. " Preferential trading arrangements as strategic positioning." *Journal of International Economics* 79: 143–59.

[47] Suwanprasert, W., 2017. " A note on jobs, jobs, jobs: a " new" perspective on protectionism of Costinot (2009)." *Economics Letters*, 157, pp.163-166.

[48] Stoyanov, Andrey and Halis M. Yildiz, 2015. " Preferential versus multilateral trade liberalization and the role of political economy." *European Economic Review* 80: 140-164.

[49] Wong, Woan Foong, 2017. " The free rider effect and market power in trade agreements." Working Paper.

[50] WTO, 2011. *World Trade Report 2011: The WTO and Preferential Trade Agreements: From Co-Existence to Coherence*." Geneva: WTO.

[51] Yi, Sang-Seung, 1996. " Endogenous formation of customs unions under imperfect competition: open regionalism is good." *Journal of International Economics* 41(1): 153-77.

Figure 1: Equilibrium FTAs under the benchmark WTO game
Figure 2: Equilibrium FTAs in the absence of the free internal trade requirement
Figure 3: Stability of global free trade and the free internal trade requirement
Figure 4: Global free trade when FTAs face a ceiling on internal tariffs
Figure 5: Global free trade under semi-cooperative MFN tariffs
Figure 6: Global free trade with three asymmetric countries

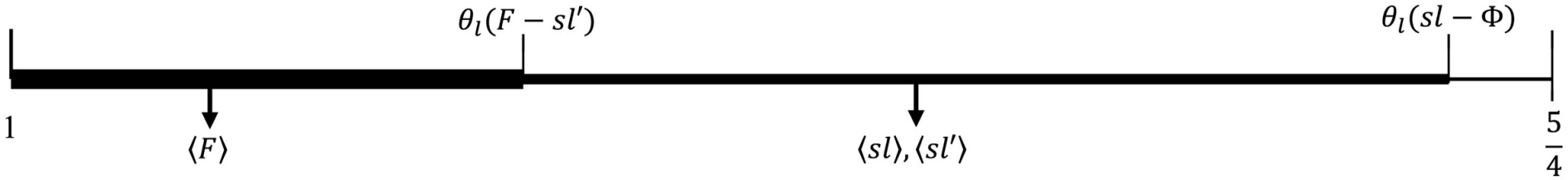
$\theta_l(F - sl')$
$\theta_l(sl - \Phi)$
1
$\langle F \rangle$
$\langle sl \rangle, \langle sl' \rangle$
$\frac{5}{4}$

Figure 1

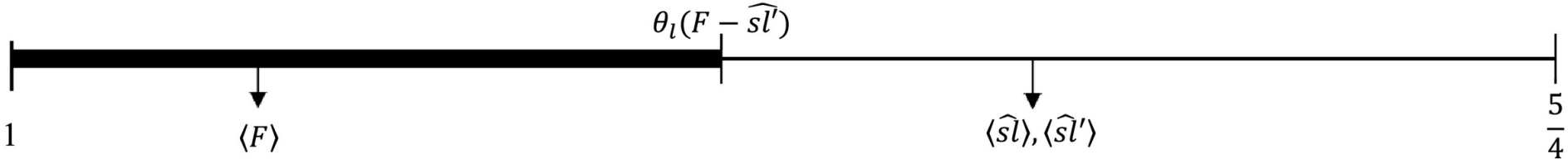
$\theta_l(F - \widehat{sl'})$
1
$\langle F \rangle$
$\langle \widehat{sl} \rangle, \langle \widehat{sl'} \rangle$
$\frac{5}{4}$

Figure 2

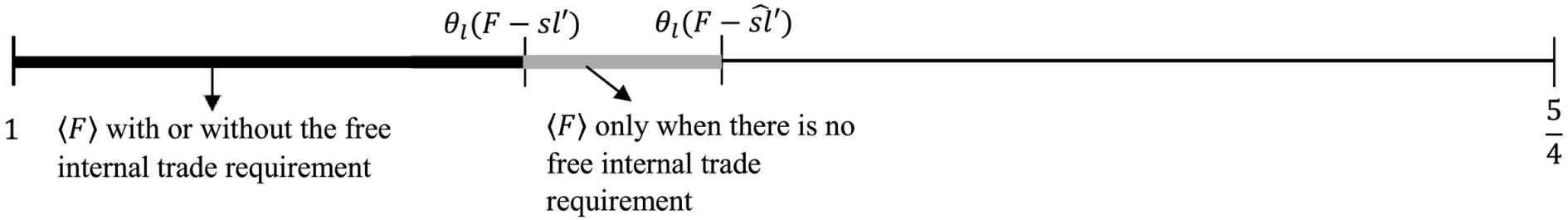
$\theta_l(F - sl')$
$\theta_l(F - \widehat{sl}')$
1
$\langle F \rangle$ with or without the free internal trade requirement
$\langle F \rangle$ only when there is no free internal trade requirement
$\frac{5}{4}$

Figure 3

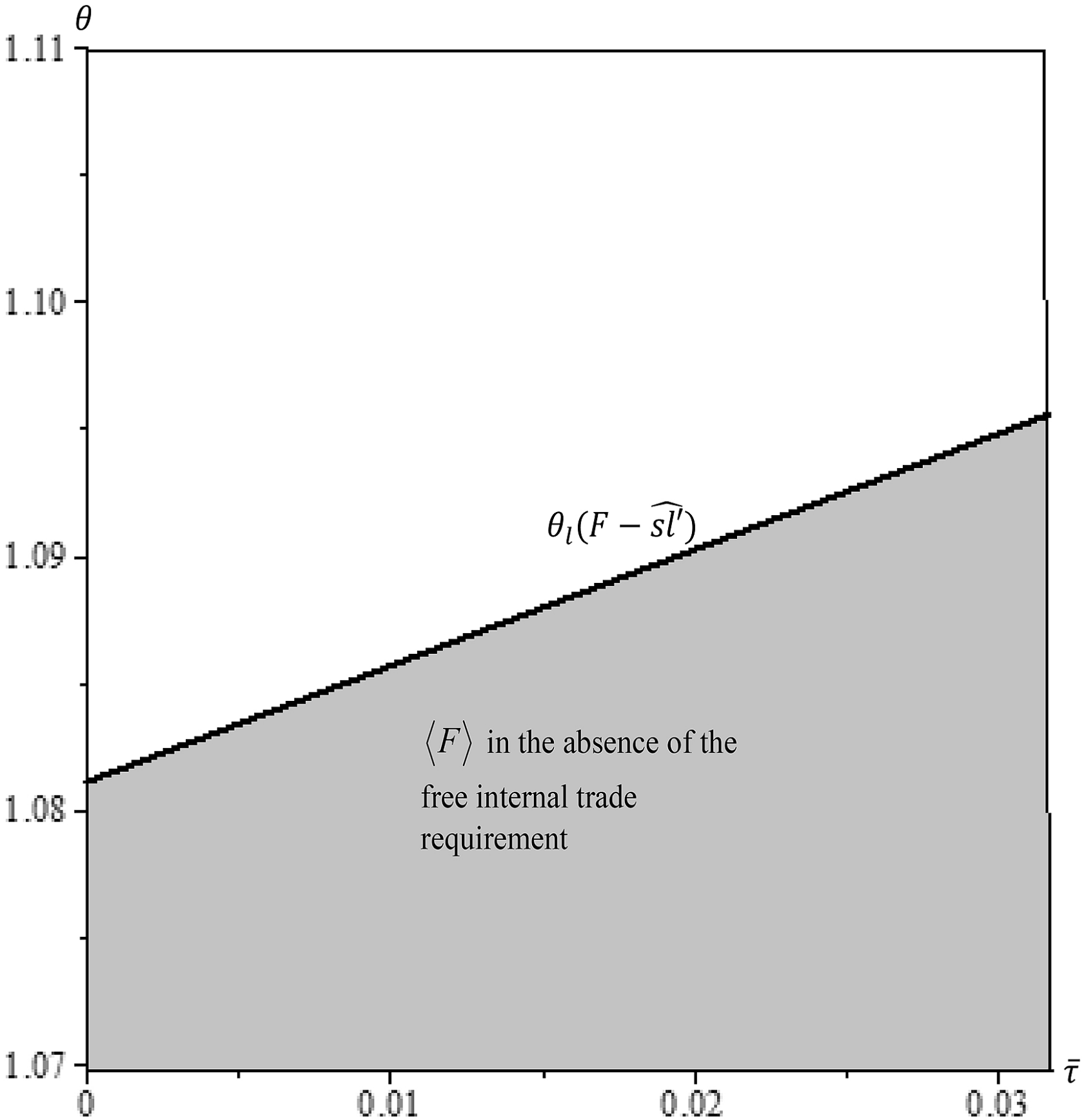
θ
1.11
1.10
1.09
1.08
1.07
0
0.01
0.02
0.03
τ̄
θ_l(F − sl′)
⟨F⟩ in the absence of the
free internal trade
requirement

Figure 4

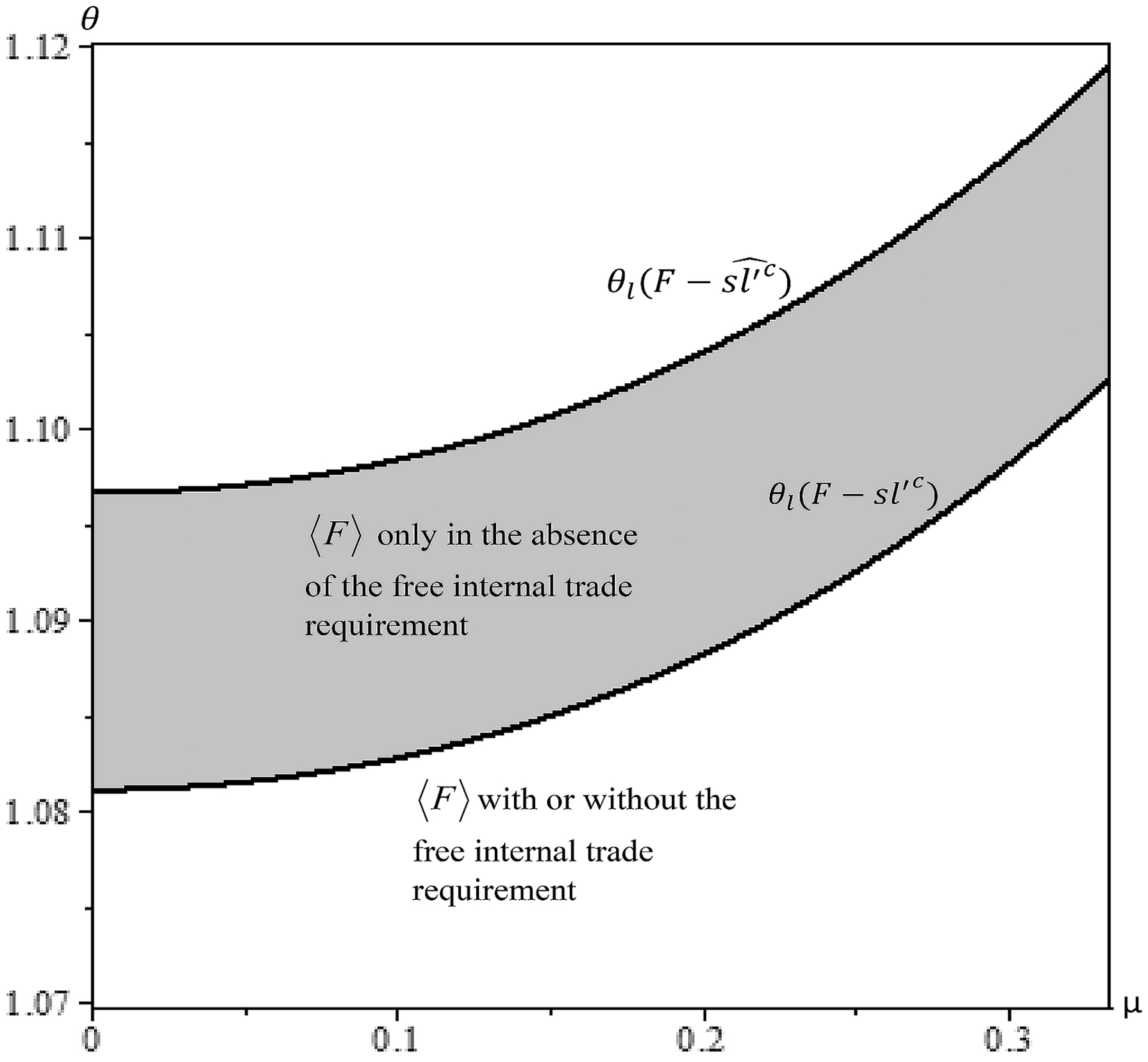
θ
1.12
1.11
1.10
1.09
1.08
1.07
0
0.1
0.2
0.3
μ
$\theta_l(F - s\widehat{l'^c})$
$\theta_l(F - sl'^c)$
$\langle F \rangle$ only in the absence of the free internal trade requirement
$\langle F \rangle$ with or without the free internal trade requirement

Figure 5

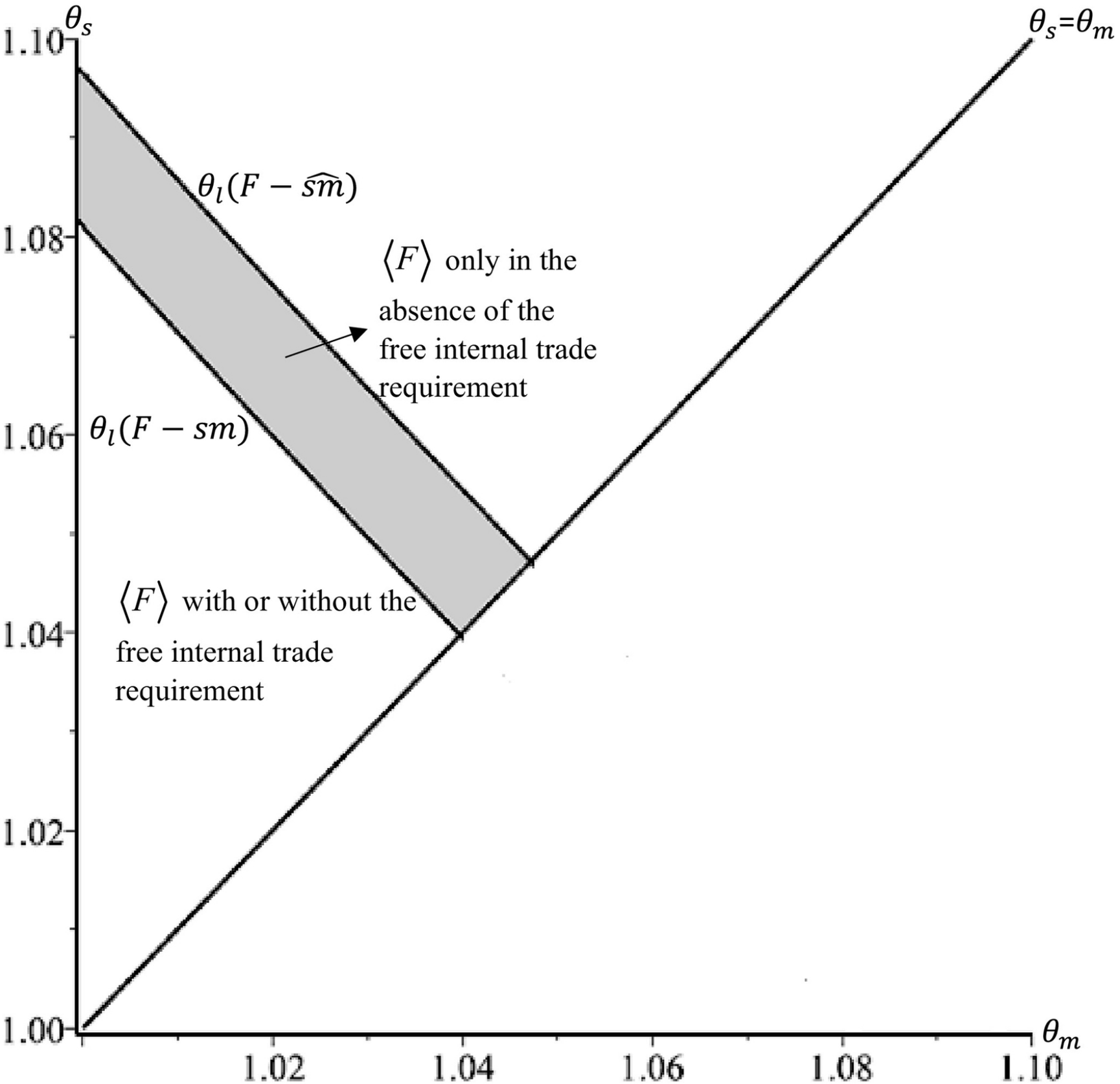
θ_s
θ_s=θ_m
1.10
1.08
1.06
1.04
1.02
1.00
$\theta_l(F - \widehat{sm})$
$\langle F \rangle$ only in the absence of the free internal trade requirement
$\theta_l(F - sm)$
$\langle F \rangle$ with or without the free internal trade requirement
θ_m
1.02
1.04
1.06
1.08
1.10

Figure 6