European University Institute

Department of Economics

EUI Working Papers

ECO 2007/17

The US Trade Deficit, the Decline of the WTO and the Rise of Regionalism

Itai Agur

EUROPEAN UNIVERSITY INSTITUTE
DEPARTMENT OF ECONOMICS

The US Trade Deficit, the Decline of the WTO and the Rise of Regionalism

ITAI AGUR

EUI Working Paper **ECO** 2007/17

ISSN 1725-6704

Printed in Italy
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy

http://www.eui.eu/
http://cadmus.eui.eu/

The US Trade Deficit, the Decline of the WTO and the Rise of Regionalism

Itai Agur* *[Job Market Paper]*

September 8, 2007

Abstract

This paper argues that the growing US trade deficit has caused the decline of the WTO and the rise of regional trade agreements. A country in deficit prefers to retain market power against countries with a large surplus. Multilateral cooperation restricts its choice. This notion is formalized in a three-country game in which countries negotiate multilaterally and, if that fails, bilaterally. The multilateral agreement only holds for sufficiently even trade balances. When one country's deficit grows too large, a regionalist equilibrium emerges. A VAR analysis shows that the US trade balance explains over 50% of the variation in regional trade agreements.

Keywords: Regionalism, RTA, Multilateralism, WTO, Trade balance, US trade deficit

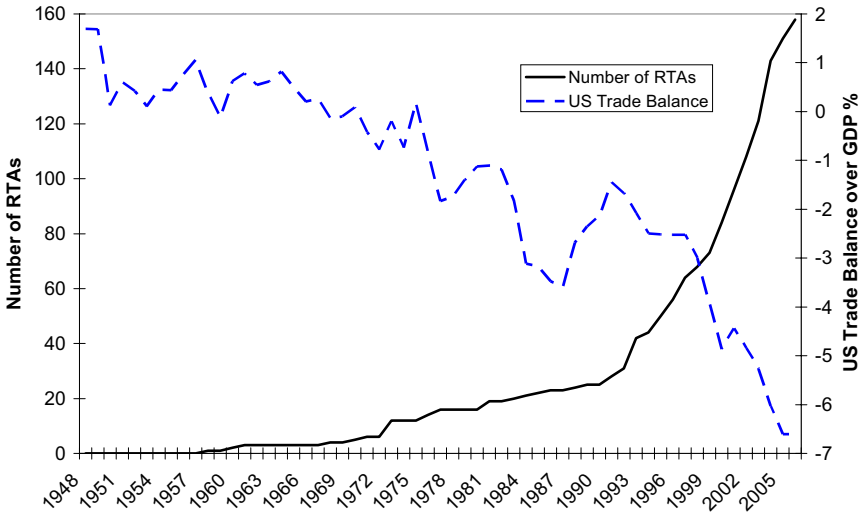
JEL Classification: C72, C73, F13, F32

*Economics Department, European University Institute, Villa San Paolo, Via della Piazzuola 43, 50133 Firenze, Italy. E-mail: itai.agur@eui.eu. I am grateful to Karl Schlag, Pascal Courty, Elhanan Helpman, Kyle Bagwell, Helmut Lűtkepohl, Renato Faccini, Philip Sauré, Liufang Chou, Sebastian Krautheim and Elie Agur for their feedback, and to Anindya Banerjee, Maria Demertzis, Michele Ruta, Ulrich Petersmann, and the participants of the EUI Trade Group and May Forum Seminars for useful discussions. All remaining errors are mine.

1 Introduction

Over the past two decades the way countries negotiate trade liberalization has changed remarkably. The multilateral process, responsible for the highly successful post-War liberalization effort, has slowed down, arguably reaching a standstill at the WTO's Doha Round. At the same time, trade agreements outside of the multilateral system have grown at an extraordinary pace. The solid line in Figure 1 depicts the number of Regional Trade Agreements (RTAs) according to WTO data.[1]

Figure 1: The number of RTAs and the US trade balance



What has made countries opt for this regional route? In what has become known as the Bhagwati-hypothesis, Bhagwati (1993, p.29) famously argued that:

> "The main driving force for regionalism today is the conversion of the United States, hitherto an abstaining party [...] The conversion of the United States is of major significance. As the key defender of multilateralism through the postwar years, its decision now to travel the regional route [...] tilts the balance of forces at the margin away from multilateralism to regionalism."

[1] This figure includes RTAs on trade in goods (www.wto.org). Note that RTAs are not always regional in nature (i.e: US-Singapore).

But what is it that made the US government suddenly change policy? Why did this happen in the 1980s and why has regionalism persisted ever since? In sum: what is the economic rationale for the hypothesis proposed by Bhagwati? This paper presents a theory capable of providing that rationale. Its explanation is based on the rise of trade imbalances. From the 1980s onwards, we argue, the growing US trade deficit has made selective trade partnerships more attractive for the US government. This has stalled multilateralism and has made countries around the world pursue the regionalist alternative. The broken line in Figure 1 displays the evolution of the US trade balance.

The basic mechanism is intuitive. When a bilateral trade imbalance is large enough, the deficit country can be better off under non-cooperation than under bilateral free trade. The reason is that a country in deficit loses much by giving up independent tariff setting. It has large market power over its import goods, and hence a strong incentive to depress their prices using import tariffs. Empirical work by Broda et al. (2006) and Bagwell and Staiger (2006) provides support for the theory that countries set higher tariffs on goods that they have more market power over. At the same time, because a deficit country exports relatively little, it gains less from lower tariffs on its exports abroad.

If a deficit country wants to cooperate at all, therefore, it will prefer to choose a partner towards which its deficit is small. It becomes selective. Instead of proceeding multilaterally, which restricts its choice, it opts for a regional route. We formalize this intuition in a negotiation game with three countries. One country, X, has a trade deficit against both others, Y and Z, but a larger deficit towards Z. One can think of X as the US, Z as Japan in the 1980s or China nowadays, and Y as the rest of the world. In an appendix we provide a microfounding model for the game's setting, where trade imbalances come about through optimal consumption smoothing in response to country-specific income shocks.

Negotiations in the game consist of two stages: multilateral and bilateral. In each stage, those agreements are signed for which all parties state that they consent. The game is solved using three properties, which we show can be derived from our microfounding goods market model: firstly, a country's benefit of a RTA increases in the trade surplus towards its partner;

secondly, there is a threshold bilateral trade deficit beyond which a country is better off without a RTA; finally, a RTA between two countries reduces the welfare of the third country.

The game has three equilibria: multilateralism, regionalism and complete non-cooperation. On the basis of the three properties, we identify a set of restrictions on the size of the bilateral trade imbalances, which relate to the equilibrium outcome. When all restrictions hold, we say that imbalances are small. When some are violated, imbalances are large. And when none hold, imbalances are extreme. For extreme imbalances there is complete non-cooperation. For large imbalances, instead, a regionalist equilibrium emerges with bilateral cooperation between $X-Y$ or $Y-Z$ or both. Multilateralism only comes about for small imbalances. Thus, rising trade imbalances can trigger the move from multilateralism to regionalism. This is the central result of the paper.

Furthermore, we analyze the welfare implications of regionalism in the presence of trade imbalances. Without the regionalist option, countries are left with a choice between multilateralism and complete non-cooperation. This restricted choice can force the deficit country into multilateral cooperation, raising world welfare. But, for sufficiently large trade imbalances, regionalism safeguards partial cooperation over complete non-cooperation. Therefore, the welfare effects of regionalism depend on the size of trade imbalances.

So far, the analysis has been purely static. However, trade balances are bound to change over time. We extend our game to a dynamic setting with recurring negotiations and a mean-reverting stochastic process for trade balances. Regionalism now always converges to world free trade in the long run. Nonetheless, the speed of convergence is slower, the larger the size of initial trade imbalances is.

Our theory leads to two testable implications. Firstly, it identifies the US trade deficit as the variable behind the Bhagwati-hypothesis. That is, the US trade balance affects multilateralism and, through it, worldwide regionalism. In terms of our game: the size of country X's deficits against Y and Z affects not only X's own partner choice, but also the presence of regionalism between Y and Z. Using a Granger-causality test we assess whether past movements in the US trade balance have predictive power over the number of RTAs signed worldwide.

We find that they do. Variance decompositions indicate, moreover, that movements in the US trade balance account for more than half of the variation of RTA growth within a 5-year horizon.

Our second testable implication is that countries select RTA partners on the basis of their bilateral trade balance. In a cross-sectional probit regression with 131 countries, Holmes (2005) finds that a larger bilateral trade imbalance implies a significantly smaller probability that two countries form a RTA. Likewise, in a panel probit Magee (2003) estimates the effect of increasing a country's bilateral trade surplus with another country from zero to ten percent of total bilateral trade in 1980. He finds that this reduces the probability that the two countries sign a RTA before 1998 by 3.4%.[2]

The next section reviews the related literature. Section 3 presents our one-shot game. Section 4 considers its dynamic extensions. Subsequently, section 5 presents the empirics. Finally, section 6 discusses the policy implications.

2 Literature review

There exist several alternatives to our explanation of the decline of multilateralism and the rise of regionalism.[3] Krugman (1993) has argued that, firstly, regionalism can go deeper than multilateralism on non-tariff issues and that, secondly, growing GATT/WTO membership has aggravated free-rider problems. Both arguments are often heard in policy debates on regionalism and the WTO (Collier (2006)). Baldwin (1997) has pointed out, however, that the regionalism of the 80s and 90s was actually primarily concerned with tariff liberalization, while Ludema (1991) has formally shown that free riding is not able to undermine multilateralism conducted under the Most Favored Nation principle.

Baldwin (1995) offers a different perspective. In his model governments only join a RTA

[2]Both Holmes (2005) and Magee (2003) interpret their results in terms of unbalanced political lobbying incentives (Grossman and Helpman (1995)). Our model, instead, would predict that their results stem from the deficit country's unwillingness to give up the ability to influence its import prices.

[3]Most of the literature on RTAs takes regionalism as given and considers its welfare implications. This literature is too extensive to review here. We refer to Winters (1996), Bhagwati et al. (1998), Panagariya (2000) and Baldwin (2005) for surveys.

if the lobbying support of exporters exceeds that of importers. When, exogenously, a single RTA is signed among two countries, lobbying efforts in other countries tilt favorably towards supporting RTA. Thus, a single event of regionalism can lead to a domino effect, which can explain why existing blocs expand. But the model does not provide a rationale for the initial choice for regionalism, nor for the slowing of the multilateral process.

Freund (2000a) does model the choice for regionalism. She shows that RTAs become more attractive when multilateral tariffs are low. For high multilateral tariffs overall efficiency incentives dominate, and non-discriminatory liberalization is best. But for low tariffs, incentives to divert trade through exclusionary RTAs gain importance. Therefore, gradual multilateral liberalization can lead to growing regionalism. Multilateralism itself remains exogenous, however.

Our paper also relates to Bagwell and Staiger (1990) and Horn et al. (2006). They show, in different settings, that when two countries cooperate on trade policy, times of trade imbalances should be paired with higher cooperative tariffs. This compensates the deficit country for the inability to make use of its market power against the other country.

3 One-shot game

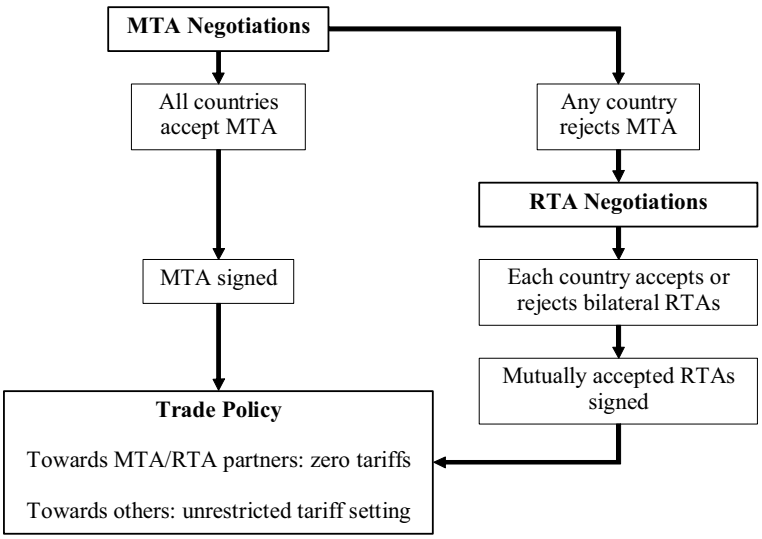
The way trade imbalances affect regionalism is best understood in a simple one-shot game. There are three players, namely countries X, Y and Z. These countries differ in their bilateral trade balances towards each other. We let $TB_{ij}(\Omega)$ denote the bilateral trade balance of country i with country j when the set of agreements Ω is in force (specified below). Here, clearly, one country's bilateral surplus is the other's bilateral deficit, $TB_{ij}(\Omega) = -TB_{ji}(\Omega)$, and world trade is always balanced, $\sum_i \sum_j TB_{ij}(\Omega) = 0$.

We obtain a simple analytical setting by ranking countries on the basis of their initial trade balances. That is, their trade balances in the absence of any agreements, $\Omega = \{\emptyset\}$. In particular, we let X be the deficit country, Y the middle country and Z the surplus country. Country X is in deficit against both others, but with a larger deficit against Z:

$TB_{XZ}(\Omega)|_{\Omega=\{\emptyset\}} < TB_{XY}(\Omega)|_{\Omega=\{\emptyset\}} < 0$. And country Z in surplus against both others, with a larger surplus against X: $TB_{XZ}(\Omega)|_{\Omega=\{\emptyset\}} < TB_{YZ}(\Omega)|_{\Omega=\{\emptyset\}} < 0$. In the appendix we present a goods market model that provides microfoundations for the setting we assume here. This microfounding model is based on two periods. In it, the game described below takes place in the second period, against given trade imbalances, which are generated through first-period income shocks.

Countries X, Y and Z play the trade negotiation game depicted in Figure 2. Thus, negotiations take the form of a statement game. First, each country announces whether it accepts or rejects the Multilateral Trade Agreement (MTA). When all accept it, the MTA is signed. But if any country rejects it, the MTA fails and RTA negotiations begin. Regionalism is thus modelled as the outside option to multilateralism. Likewise, during RTA negotiations only those agreements are signed on which there is mutual consent. As a tie-breaking assumption, countries accept an agreement towards which they are indifferent. We denote the RTA between countries i and j by RTA_{ij}. Overall, at the end of negotiations the set of agreement, Ω, can include either the MTA, or any of RTA_{XY}, RTA_{XZ}, RTA_{YZ}, or it can remain the empty set $\Omega = \{\emptyset\}$.[4]

Figure 2: Structure of the Game



[4]Saggi and Yildiz (2006) and Zissimos (2007 forthcoming) also model RTA negotiations as a statement game. For an alternative approach, in which a lead country decides whom to offer an agreement to, see Aghion et al. (2007).

After negotiations have been completed, tariffs are set. Countries implement zero tariffs against their MTA/RTA partners. We assume that countries commit themselves to the agreements that they sign during negotiations, and cannot subsequently defect from them. The implications of this assumption are discussed in section 4.3. However, an agreement between two countries imposes no constraints on their policies towards the third country.[5]

Thus, we have defined the game's players and their actions. Finally, we assume that governments are social welfare maximizers, and define three properties concerning the way that trade agreements and trade balances affect a country's welfare. Our microfounding model in the appendix shows that these properties can be formally derived from a general specification with maximizing agents. Below we briefly explain the intuition behind each property.

P1 A country's net benefit of having a bilateral agreement with another country is increasing in the trade surplus towards it:

$$\frac{\partial}{\partial\, TB_{ij}\left(\Omega\right)|_{RTA_{ij}\notin\Omega}}\left[W_i|_{RTA_{ij}\in\Omega}-W_i|_{RTA_{ij}\notin\Omega}\right]>0$$

where W_i is country i's sum of producer surplus, consumer surplus and tariff revenues.

P2 For a given set of agreements (Ω) that does not include RTA_{ij} ($RTA_{ij}\notin\Omega$) there exists a threshold bilateral trade deficit, $\overline{TB}_{ij}<0$, below which country i is better off not signing RTA_{ij}. That is, country i's net benefit of the RTA is zero at the threshold:

$$W_i|_{RTA_{ij}\in\Omega}-W_i|_{RTA_{ij}\notin\Omega}=0\Longleftrightarrow TB_{ij}\left(\Omega\right)|_{RTA_{ij}\notin\Omega}=\overline{TB}_{ij}$$

P3 Two insiders to a RTA impose a negative externality on the outsider's welfare:

$$W_i|_{RTA_{j\neq i}\in\Omega}-W_i|_{RTA_{j\neq i}\notin\Omega}<0$$

[5] That is, say, $X-Y$ RTA does not impose any constraints on either X or Y's indepedent tariff setting against Z. In our game RTAs are Free Trade Areas (FTAs), therefore, as opposed to Customs Unions. In the latter members set a common external tariff. In reality 84% of all RTAs are FTAs (Crawford and Fiorentino (2005)).

and worsen the outsider's trade balance:

$$TB_{ij}|_{RTA_{j\neq i}\in\Omega} - TB_{ij}|_{RTA_{j\neq i}\notin\Omega} < 0$$

Intuitively, the deficit country has more market power over its imports from the surplus country than vice versa. Thus, giving up the ability to influence the prices of import goods is more costly for the deficit country, which is the first property. This gives rise to a trade-off between the efficiency gains of mutual tax elimination and the cost of yielding unrestricted tariff setting. Under balanced trade the bilateral trade agreement is always mutually beneficial. But beyond a threshold bilateral imbalance, the deficit country is better off without an agreement.[6] This is the intuition behind the second property. Finally, the third property stems from the fact that the two insiders to a RTA raise the demand for, and thereby the prices of, each other's export goods. Hence, the outsider country pays more for its imports from them, which lowers its welfare and reduces its bilateral trade balances.[7]

Taken together, properties P1-P3 are sufficient to solve the game depicted in Figure 2 (proof in the appendix):

Proposition 1 *There are three types of Nash Equilibria: multilateralism ($\Omega = \{MTA\}$), regionalism ($\Omega = \{RTA_{XY}\}$, $\Omega = \{RTA_{YZ}\}$ or $\Omega = \{RTA_{XY}, RTA_{YZ}\}$), and complete non-cooperation ($\Omega = \{\emptyset\}$).*

1. *Multilateralism obtains when trade imbalances are small enough that* $TB_{XY}(\Omega)|_{\Omega=\{RTA_{YZ}\}} \geq \overline{TB}_{XY}$ and $TB_{XZ}(\Omega)|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} \geq \overline{TB}_{XZ}$ *hold.*

2. *Regionalism comes about when trade imbalances are sufficiently large that at least one of the above conditions is violated, but not so large that for all i with $TB_{ij} < 0$ it holds that $TB_{ij}(\Omega) < \overline{TB}_{ij} \forall \Omega$.*

[6] This resembles Kennan and Riezman's (1988) result that a big country can win a trade war against a small country.

[7] The price effect of a RTA on outsiders is empirically documented by Chang and Winters (2002) for the case of MERCOSUR. Thus, there is evidence that even fairly small countries can influence their terms-of-trade using trade policy. See also Broda et al. (2006), Bagwell and Staiger (2006), Anderson and Van Wincoop (2001) and Olarreaga et al. (1999).

3. *When for all i with $TB_{ij} < 0$ it holds that $TB_{ij}(\Omega) < \overline{TB}_{ij} \forall \Omega$, then there is complete non-cooperation.*

This is the central result of the paper. When trade imbalances are small enough, multilateral cooperation is feasible. But when they become too large, the deficit country prefers to choose its partner selectively. This country's decision to opt out of multilateralism subsequently gives rise to worldwide regionalism.

But does this mean that world welfare is reduced by the existence of the regionalist option? Assume, as is standard in international trade theory, that world welfare is largest under world free trade. The welfare effect of regionalism then depends on the size of trade imbalances. For moderate imbalances regionalism is damaging. Consider the case where X loses a little from cooperating with Z, but gains much from a RTA with Y. Then, if X has the regionalist alternative, it rejects world free trade. But given the large gains from RTA_{XY}, X prefers the MTA to complete non-cooperation. That is: restricting X's choice to multilateralism or complete non-cooperation, it opts for multilateralism.

Now consider the opposite: X has a large trade deficit towards Z and gains much from retaining independent tariff setting against it, while X is indifferent towards a RTA with Y. When X is faced with the choice between multilateralism and non-cooperation only, it chooses non-cooperation. In this case, regionalism safeguards partial free trade.[8] Formally (proof in the appendix):

Proposition 2 *Consider a No-Regionalism Game, in which rejection of the MTA implies complete non-cooperation, $\Omega = \{\emptyset\}$.*

1. *Whenever $W_X|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} > W_X|_{\Omega=\{MTA\}} \geq W_X|_{\Omega=\{\emptyset\}}$ multilateralism holds in the No-Regionalism Game but not in the standard Game.*

2. *Whenever $W_X|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} \geq W_X|_{\Omega=\{\emptyset\}} > W_X|_{\Omega=\{MTA\}}$ complete non-cooperation comes about in the No-Regionalism Game but not in the standard Game.*

[8] Absent trade diversion (as in our microfoundations), partial cooperation is certainly better for world welfare than non-cooperation.

Trade imbalances are a distortion. They create harmful asymmetries between countries. The first best solution would be the elimination of imbalances, or the establishment of a system of international transfers. When this is impossible, regionalism can sometimes be a second best solution.

4 Dynamic extensions

The results we have derived so far are based on a purely static setting. Yet, trade balances change over time. This section presents repeated games in which the interaction over time between trade imbalances and trade negotiations is analyzed. We start from a basic repeated game, where trade imbalances arise only once, and vanish over time. Section 4.2 extends to a richer, more realistic dynamic process with repeated shocks and mean reversion. Unlike for our one-shot game, however, we do not provide microfoundations for these settings. The evolution of trade balances is exogenously specified. In particular, the game in section 4.1 presents a dynamic process that does not have the potential to be microfounded, as it would violate long-run debt repayment. However, it should be seen as an initial example that allows us to present the main intuitions of the dynamic setting in a clear manner. These intuitions are subsequently of use in understanding the mechanisms of the dynamic game with repeated shocks.

4.1 Dynamic game with temporary imbalances

In the dynamic game each period is a precise repetition of the one-shot game in Figure 2, except that negotiations only take place about agreements that have not previously been signed. Moreover, if a MTA is signed, it supplants any existing RTAs. Governments discount future periods at rate $\delta \in [0, 1]$.

Consider the following development of trade balances. In the first period initial trade balances are as in the one-shot game: $TB_{XZ}^{t=1}(\Omega^{t=1})|_{\Omega^{t=1}=\{\emptyset\}} < TB_{XY}^{t=1}(\Omega^{t=1})|_{\Omega^{t=1}=\{\emptyset\}} < 0$ and $TB_{XZ}^{t=1}(\Omega^{t=1})|_{\Omega^{t=1}=\{\emptyset\}} < TB_{YZ}^{t=1}(\Omega^{t=1})|_{\Omega^{t=1}=\{\emptyset\}} < 0$. However, balanced trade is restored in

the long run: $\lim_{t\to\infty} TB_{ij}^{t}\left(\Omega^{t}\right) = 0 \forall i, j, \Omega$.

In particular, between $t = 1$ and the limit $t \to \infty$ there is a smooth decay of trade imbalances. This allows us to focus in a simple way on the question whether temporary imbalances can lead to permanent regionalism. We let trade imbalances decay by a factor $\rho \in (0, 1)$. For each country pair i,j we have that: $TB_{ij}^{t}\left(\Omega^{t}\right)\big|_{\Omega^{t}=\Omega^{t-1}} = \rho TB_{ij}^{t-1}\left(\Omega^{t-1}\right)$. That is, at the beginning of period t, before new negotiations take place ($\Omega^{t} = \Omega^{t-1}$) each trade balance is a fraction ρ of its value at the end of period $t-1$. For instance, if $\rho = 0.9$ and at the end of a period $TB_{XZ} = -10$ ($TB_{ZX} = 10$), then at the beginning of the next period: $TB_{XZ} = -9$ ($TB_{ZX} = 9$).

We solve the game by considering first the decisions taken in the limit for different structures of previously signed agreements. The decisions in the first period can then be derived by backward induction. The outcome is as follows (proof in the appendix):

Proposition 3 *There are three equilibrium paths to the dynamic game: multilateralism forever; static regionalism with one RTA forever; expanding regionalism, converging to world free trade. We call these the Good, the Bad and the Ugly Paths, respectively:*

1. *If* $\lim_{t\to\infty}\left[W_{i}^{t}|_{\Omega=\{RTA_{ij}\}} - W_{i}^{t}|_{\Omega=\{MTA\}}\right] \leq 0$ *then for small first period imbalances (as in point 1 of Proposition 1) the Good Path obtains, and otherwise the Ugly Path.*

2. *If* $\lim_{t\to\infty}\left[W_{i}^{t}|_{\Omega=\{RTA_{ij}\}} - W_{i}^{t}|_{\Omega=\{MTA\}}\right] > 0$ *then there exists a value* $\overline{\delta}$ *of the discount rate such that for* $\delta \geq \overline{\delta}$ *all three paths are possible, and for* $\delta < \overline{\delta}$ *paths are as in (1).*

Here, $W_{i}^{t}|_{\Omega=\{RTA_{ij}\}} > W_{i}^{t}|_{\Omega=\{MTA\}}$ means that even under balanced trade ($t \to \infty$) a country is better off being the insider to the only RTA than being in a MTA. Notice that properties P1-P3 do not exclude this possibility. Intuitively, RTA insiders can gain rents by improving each other's terms-of-trade at the outsider's expense. The MTA destroys such rents.

Proposition 3 shows that, like in the one-shot game, trade imbalances can trigger regionalism. When that regionalism is primarily driven by the benefits partners impose on each other, rather than their gains from excluding others, it always follows the expanding path. We call

this the Ugly Path, because there is a growing web of RTAs. Eventually, that web is replaced by a MTA.

However, when the gains from excluding the outsider are large, there is also the potential for static regionalism that does not achieve world free trade in the long run. We call this the Bad Path. When there is only one RTA and trade is balanced, RTA insiders face a trade-off. On the one hand, an insider has the incentive to sign an agreement with the excluded country. In this manner, it would become the insider on two RTAs and gain instantaneous welfare. On the other hand, it knows that the next period, the outsiders to the two respective agreements will be better off signing the last RTA and achieving world free trade. Due to the size of RTA rents, world free trade is worse for the insider than one RTA. When patience is high enough future losses outweigh instantaneous gains, and regionalism that was triggered by temporary imbalances may perpetuate.

4.2 Dynamic game with repeated shocks

The game above provides the main intuitions of the dynamic setting. In reality, however, trade balances are continuously subject to shocks. Extending the game, we propose the following dynamics: $TB_{ij}^{t}\left(\Omega^{t}\right)\big|_{\Omega^{t}=\Omega^{t-1}} = \rho TB_{ij}^{t-1}\left(\Omega^{t-1}\right) + \varepsilon_{ij}^{t}$ for each country pair i,j. And, by $TB_{ij}^{t}\left(\Omega^{t}\right) = -TB_{ji}^{t}\left(\Omega^{t}\right)$, this implies: $TB_{ji}^{t}\left(\Omega^{t}\right)\big|_{\Omega^{t}=\Omega^{t-1}} = \rho TB_{ji}^{t-1}\left(\Omega^{t-1}\right) - \varepsilon_{ij}^{t}$. Thus, the difference with section 4.1 is the presence of shocks, ε_{ij}^{t}. These shocks are realized at the beginning of each period, prior to the negotiations depicted in Figure 2.

Shocks have zero mean and are uncorrelated across different country pairs and over time: $E\left[\varepsilon_{ij}^{t}\right] = 0$, $E\left[\varepsilon_{ij}^{t}\varepsilon_{ij}^{t-p}\right] = 0 \forall p$, $E\left[\varepsilon_{XY}^{t}\varepsilon_{XZ}^{t}\right] = 0$, $E\left[\varepsilon_{XY}^{t}\varepsilon_{YZ}^{t}\right] = 0$, and $E\left[\varepsilon_{XZ}^{t}\varepsilon_{YZ}^{t}\right] = 0$. Given $\rho > 0$, shocks persist over time. But by $\rho < 1$ the process is mean reverting in the long run (where the mean is balanced bilateral trade).

We solve the game by showing that neither complete non-cooperation, nor a single RTA, nor two RTAs can be sustained forever. Therefore, world free trade is the unique long run equilibrium (proof in appendix):

Proposition 4 *World free trade is always achieved in the long run:* $\lim_{t\to\infty} \Omega^t = \{MTA\}$. *But convergence to world free trade is expected to take longer the larger are first period trade imbalances,* $\left|TB_{ij}^{t=1}\left(\Omega^{t=1}\right)\right|$.

Thus, regionalism now unambiguously converges to world free trade in the long run. The size of trade imbalances only affects the time it takes to achieve world free trade. Contrary to the dynamic game with temporary imbalances, therefore, static regionalism cannot be permanently sustained. At some point the incentives of partners to a rent-extracting RTA diverge, and their exclusionary cooperation ends.

Proposition 4 is a sensitive result in two respects, however. Firstly, its proof relies on the infinite repetition of the game, and cannot be simply extended to finite periods. Secondly, the assumption that countries commit themselves to agreements forever plays an important role, as discussed below.

4.3 The commitment assumption

In the literature on RTAs there are two strands regarding commitment. One which, like us, assumes the ability to commit, and another which requires agreements to be self-enforcing mechanisms.[9] The incentive to deviate from trade agreements arises from a Prisoner's Dilemma type of structure. After all, each country has the incentive to raise its own welfare at others' expense, by depressing the world prices of its import goods with tariffs (Bagwell and Staiger, 1999).

However, the commitment strand considers that, in reality, deviation from trade agreements can give only a very short-lived benefit to the defector, if any. Generally, trade measures are quickly observable to other countries. In fact, in our game one period equals one multilateral negotiation round. Such rounds can be more than a decade apart. Deviation profit would constitute only a fraction of the "period" income, therefore. The other strand takes the

[9] The first strand includes Grossman and Helpman (1995), Levy (1997), Burbidge et al. (1997), Krishna (1998), Bagwell and Staiger (1999), McLaren (2002), Maggi and Rodriguez-Clare (1998, 2006), and Ornelas (2005). The latter includes Riezman (1991), Bagwell and Staiger (1997a, 1997b), Freund (2000b), Bond et al. (2001), Conconi and Perroni (2003) and Saggi (2006).

view that as there is no institution that punishes defectors, agreements must necessarily be self-enforcing.

Without commitment, a one-shot game has the same outcome as a one period Prisoner's Dilemma game. Furthermore, in a dynamic game cooperative outcomes are only sustainable for sufficiently high discount rates (Riezman, 1991). However, in our dynamic game with repeated shocks even high patience does not ensure cooperation. At a point in time when there is balanced trade, countries have the highest expected welfare from world free trade forever, and can agree to a MTA. But after a shock in the future, one of the countries can have higher expected welfare from abandoning the agreement. Without commitment, therefore, multilateral cooperation would be unstable even when governments are patient. Temporary bouts of regionalism would be a permanent feature of the world trading system.

5 Empirics

Relating back to the Bhagwati-hypothesis discussed in our introduction, our theory identifies the US trade balance as the key variable driving the decline of multilateralism and the rise of regionalism. In terms of our game, a growing trade deficit of country X brings multilateralism to a halt. By doing so, it triggers not only its own regionalism, but also regionalism between other countries: countries Y and Z may sign a RTA with each other when the MTA fails.

This section tests for Granger-causality between the US trade balance and worldwide regionalism. Granger-causality means that past movements in one variable have explanatory power over movements in another variable (Lűtkepohl (2005)). If our theory is correct, we would expect to find that past movements in the US trade balance precede, that is, Granger-cause, the growth of RTAs.

We have two variables. The first is the number of RTAs, obtained from www.wto.org. The WTO reports RTA data, splitting agreements into RTAs on goods and RTAs on services. We use only the RTAs on goods, since our theory is not concerned with agreements on services. Moreover, the WTO adjusts its data for inactive agreements. For instance, many EU entrants

already had a RTA with the EU bloc prior to entering. Such agreements become inactive upon EU accession. The WTO database then counts the first RTA as an agreement in the year that it is signed, and upon EU accession counts the accession as an agreement, while deleting the previous RTA. In a time-series sense, the move from RTA to accession is a zero entry (plus one, minus one).

The RTA variable is, of course, a discrete variable. In principle, the Granger-causality test is designed for continuous variables. However, Figure 1 shows that the RTA variable does not exhibit discrete jumps. It grows smoothly from 0 to 158. We thus treat it as a continuous variable. The second variable is the US trade balance over GDP, taken from IMF-IFS. We focus on the last thirty years, as the rise of the US trade deficit is a fairly recent phenomenon. And, so long as imbalances remain small enough that no thresholds are crossed, our theory predicts no relationship between variation in imbalances and in regionalism (Proposition 1). However, for robustness we also report results for the longest attainable sample, 1948-2006.

Importantly, both variables are non-stationary. Adjusted Dickey-Fuller tests indicate that the trade balance is I(1) while the number of RTAs is I(2). This is true for both the past thirty years and the entire 1948-2006 sample. Lűtkepohl (2005, p. 320) and Lűtkepohl and Reimers (1992) argue that Granger-causality based on an ordinary VAR is not spurious when two I(1) variables are cointegrated. Our first steps, therefore, are to make both variables of the same order of integration and, subsequently, test for cointegration. We take the first differences of the RTA variable, from which an I(1) variable representing RTA growth is obtained.

A Johansen Trace test indicates that the two variables indeed have cointegration rank 1. This is true for lag lengths suggested by both AIC and Schwarz criteria, and is robust to the inclusion of a deterministic trend. Moreover, this holds for both the past thirty years and the 1948-2006 sample. We can thus proceed in the way described by Lűtkepohl (2005) and Lűtkepohl and Reimers (1992). The VAR takes the form:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} + \sum_{i=1}^{p} \begin{bmatrix} \beta_{11,i} & \beta_{12,i} \\ \beta_{21,i} & \beta_{22,i} \end{bmatrix} \begin{bmatrix} y_{1,t-i} \\ y_{2,t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \tag{1}$$

where y_1 and y_2 are the two variables, C_1 and C_2 are constants, p is the chosen lag length and u_1 and u_2 are white-noise error terms. Then, y_{1t} is not Granger-causal for y_{2t} if and only if

$$\beta_{21,i} = 0 \forall i \in [0, p] \tag{2}$$

Using a Wald test statistic this null hypothesis is tested against the alternative that at least one of the $\beta_{21,i}$ is nonzero.

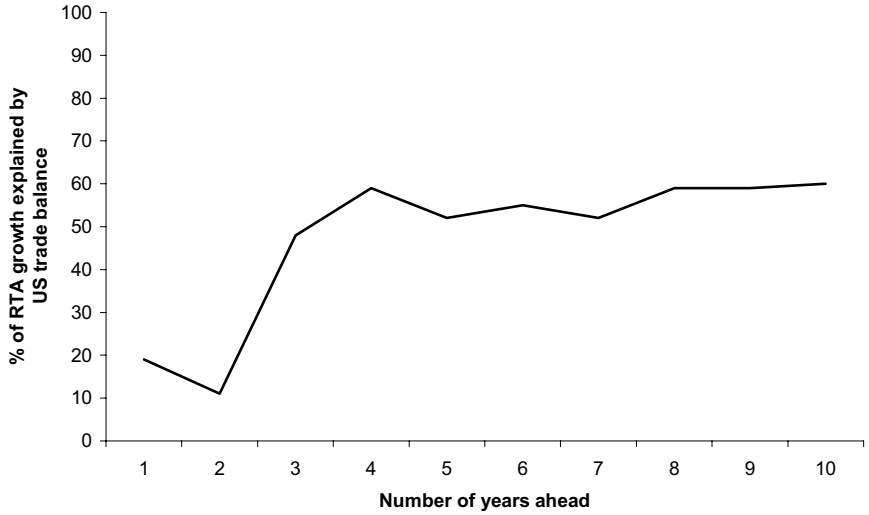
For the past thirty years, allowing up to 10 lags both the AIC and Schwarz criteria select 10 lags. A long lag length is in line with theoretical expectations: much time can go between a trade-balance induced policy shift, and completion of RTA negotiations. However, for 10 lags there are insufficient observations left to conduct a Granger-causality test. We then allow up to 9 lags, and both measures select 9 lags. We now find that Granger-causality runs only in the direction predicted by the model. Granger-noncausality from the trade balance to RTA growth is rejected at 5% significance, with a p-value of 0.039. Granger-noncausality in the opposite direction is not rejected, the p-value being 0.375.

For the entire 1948-2006 sample, however, both AIC and Schwarz criteria select only one lag. Granger-noncausality from the trade balance to RTA growth is rejected, with a p-value of 0.005. However, Granger-noncausality in the opposite direction is also rejected, p-value 0.008.

To give a richer quantitative picture, we also perform a variance decomposition. A variance decomposition shows, for different time horizons, how much of the variation in one variable is explained by the variation in the other. This is a standard methodology for interpreting VARs (Lűtkepohl (2005)). Figure 3 portrays the results for the past thirty years.

In Figure 3, movements in the US trade balance explain as much as 60% of the variation in RTA growth within a 4-year horizon. For the entire 1948-2006 sample, however, this is only 11%. Thus, the relationship between the two variables seems to have fundamentally changed from the first to the second half of the 1948-2006 sample. Again, this makes sense when viewed from our theory. Until thresholds are crossed there is no relationship between trade imbalances and regionalism (Proposition 1).

Figure 3: RTA Growth - Variance Decomposition, 1976-2006



6 Policy implications

Our theory relates the policy debate on global trade imbalances to the policy debate on the WTO. Until now these have been largely separate. If anything, some policy makers have suggested a role for the WTO in helping to bring down global imbalances (Dodge (2005), De Rato (2005)). Our work indicates the importance of the opposite channel: limiting imbalances to garner incentives for faster trade liberalization. Thus, the future success of the WTO may hinge upon such factors as the US government's willingness to cut its budget deficit and the Chinese government's exchange rate policy.

Our theory also relates to the policy question whether GATT Article XXIV should be strengthened. This Article governs the exception to non-discriminatory (multilateral) liberalization granted to RTAs. Proponents of strengthening this Article (Bhagwati (1993)) argue that larger barriers to RTA formation could aid world trade liberalization through the multilateral process. Our theory is in agreement with this statement only when trade imbalances are not too large. Otherwise, no trade cooperation takes place at all in the absence of regionalism.

A Proofs

Proof of Proposition 1. Applying backward induction, we consider first the outcome of RTA negotiations and then that of MTA negotiations. Below is country X's payoff matrix during RTA negotiations:

X's actions:	Other players' actions:	
	$RTA_{YZ} \in \Omega$	$RTA_{YZ} \notin \Omega$
Reject both	$W_X\vert_{\Omega=\{RTA_{YZ}\}}$	$W_X\vert_{\Omega=\{\emptyset\}}$
Accept only RTA_{XY}	$W_X\vert_{\Omega=\{RTA_{XY},RTA_{YZ}\}}$	$W_X\vert_{\Omega=\{RTA_{XY}\}}$
Accept only RTA_{XZ}	$W_X\vert_{\Omega=\{RTA_{XZ},RTA_{YZ}\}}$	$W_X\vert_{\Omega=\{RTA_{XZ}\}}$
Accept both	$W_X\vert_{\Omega=\{RTA_{XY},RTA_{XZ},RTA_{YZ}\}}$	$W_X\vert_{\Omega=\{RTA_{XY},RTA_{XZ}\}}$

where it is implicit that whenever X accepts RTA_{XY} or RTA_{XZ}, Y or Z accept it too. This follows from P1 and P2: $TB_{ij}(\Omega) > 0 > \overline{TB}_{ij}$. While by initial imbalances, $TB_{XZ}(\Omega)|_{\Omega=\{\emptyset\}} < TB_{XY}(\Omega)|_{\Omega=\{\emptyset\}} < 0$ and from P3, $TB_{XY}|_{RTA_{YZ}\in\Omega} - TB_{XY}|_{RTA_{YZ}\notin\Omega} < 0$ and $TB_{XZ}|_{RTA_{YZ}\in\Omega} - TB_{XZ}|_{RTA_{YZ}\notin\Omega} < 0$ country X is in bilateral deficit against Y and Z regardless of the actions of Y and Z against each other. Thus, irrespective of whether $RTA_{YZ} \in \Omega$ or $RTA_{YZ} \notin \Omega$, the RTAs that country X accepts are always signed.

Now, consider the case when $TB_{XZ}(\Omega)|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} \geq \overline{TB}_{XZ}$ and $TB_{XY}(\Omega)|_{\Omega=\{RTA_{YZ}\}} \geq \overline{TB}_{XY}$ hold (point 1 in Proposition 1). P2 in conjunction with $TB_{XZ}(\Omega)|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} \geq \overline{TB}_{XZ}$ gives

$$W_X|_{\Omega=\{RTA_{XY},RTA_{XZ},RTA_{YZ}\}} > W_X|_{\Omega=\{RTA_{XY},RTA_{YZ}\}}$$

which, in turn, implies that

$$W_X|_{\Omega=\{RTA_{XY},RTA_{XZ},RTA_{YZ}\}} > W_X|_{\Omega=\{RTA_{XZ},RTA_{YZ}\}}$$

Likewise, P2 together with $TB_{XY}(\Omega)|_{\Omega=\{RTA_{YZ}\}} \geq \overline{TB}_{XY}$ gives

$$W_X|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} > W_X|_{\Omega=\{RTA_{YZ}\}}$$

Hence, when Y and Z accept RTA_{YZ}, then $\Omega = \{RTA_{XY}, RTA_{XZ}, RTA_{YZ}\}$ achieves maximum welfare for country X and the action "Accept both" is optimal. However, by P3, the above welfare rankings directly imply

$$W_X|_{\Omega=\{RTA_{XY},RTA_{XZ}\}} > W_X|_{\Omega=\{RTA_{XY}\}}$$

$$W_X|_{\Omega=\{RTA_{XY},RTA_{XZ}\}} > W_X|_{\Omega=\{RTA_{XZ}\}}$$

$$W_X|_{\Omega=\{RTA_{XY}\}} > W_X|_{\Omega=\{\emptyset\}}$$

Therefore, "Accept both" is X's dominant strategy (i.e. optimal regardless of whether Y and Z accept RTA_{YZ}).

Moreover, whenever X is willing to accept RTA_{XZ} then, by $TB_{XZ}(\Omega)|_{\Omega=\{\emptyset\}} < TB_{YZ}(\Omega)|_{\Omega=\{\emptyset\}} < 0$, Y is also willing to accept RTA_{YZ} (as is Z, by P2). Hence, during RTA negotiations all RTAs are accepted. Then, during the preceding MTA negotiations, countries face the choice between $\Omega = \{MTA\}$ if all accept the MTA and $\Omega = \{RTA_{XY,}RTA_{XZ}, RTA_{YZ}\}$ if the MTA is rejected. Since both imply world free trade $W_i|_{\Omega=\{RTA_{XY,}RTA_{XZ},RTA_{YZ}\}} = W_i|_{\Omega=\{MTA\}}\ \forall i$ and by the tie-breaking assumption the MTA is accepted by all. This proves point 1 in Proposition 1.

Next, consider the case that for all i with $TB_{ij} < 0$ it holds that $TB_{ij}(\Omega) < \overline{TB}_{ij}\forall\Omega$ (point 3 in Proposition 1). During RTA negotiations for each country pair there is one country (i with $TB_{ij} < 0$) that rejects the RTA, regardless of other countries' actions ($\forall\Omega$). Hence, RTA negotiations have $\Omega = \{\emptyset\}$ as the outcome. At the preceding MTA negotiations, countries thus choose between $\Omega = \{MTA\}$ and $\Omega = \{\emptyset\}$. Here, country X certainly rejects MTA. This follows from the fact that $TB_{Xj}(\Omega) < \overline{TB}_{Xj}\forall\Omega$ implies

$$W_X|_{\Omega=\{\emptyset\}} > W_X|_{\Omega=\{RTA_{XY},RTA_{XZ}\}}$$

and, by P3,

$$W_X|_{\Omega=\{RTA_{XY},RTA_{XZ}\}} > W_X|_{\Omega=\{RTA_{XY},RTA_{XZ},RTA_{YZ}\}}$$

so that by $W_X|_{\Omega=\{RTA_{XY,}RTA_{XZ},RTA_{YZ}\}} = W_X|_{\Omega=\{MTA\}}$ we have

$$W_X|_{\Omega=\{\emptyset\}} > W_X|_{\Omega=\{MTA\}}$$

and country X rejects the MTA. Thus, the outcome of trade negotiations is $\Omega = \{\emptyset\}$ which proves point 3 in Proposition 1.

Finally, when trade balances do not satisfy the conditions in either point 1 or point 3 of Proposition 1, then not all RTAs are accepted, nor are all RTAs rejected. The outcome is either $\Omega = \{RTA_{XY}\}$ or $\Omega = \{RTA_{YZ}\}$ or $\Omega = \{RTA_{XY}, RTA_{YZ}\}$. By $TB_{XZ}(\Omega)|_{\Omega=\{\emptyset\}} < TB_{XY}(\Omega)|_{\Omega=\{\emptyset\}} < 0$ and $TB_{XZ}(\Omega)|_{\Omega=\{\emptyset\}} < TB_{YZ}(\Omega)|_{\Omega=\{\emptyset\}} < 0$, if any RTA is rejected, then RTA_{XZ} is rejected by country X. When RTA_{XZ} is rejected, it holds that

$$W_X|_{\Omega=\{MTA\}} = W_X|_{\Omega=\{RTA_{XY,}RTA_{XZ},RTA_{YZ}\}} < W_X|_{\Omega=\{RTA_{XY,}RTA_{YZ}\}}$$

and, by P3,

$$W_X|_{\Omega=\{RTA_{XY,}RTA_{YZ}\}} < W_X|_{\Omega=\{RTA_{XY}\}}$$

so that, regardless of whether Y and Z sign RTA_{YZ}, X achieves higher welfare without the MTA. Thus, when X rejects any RTA it also rejects the MTA.

It remains to show that all three regionalist equilibria, $\Omega = \{RTA_{XY}\}$, $\Omega = \{RTA_{YZ}\}$ and $\Omega = \{RTA_{XY}, RTA_{YZ}\}$, exist. Three examples suffice. Take $TB_{XZ}(\Omega)|_{\Omega=\{RTA_{XY}\}} < \overline{TB}_{XZ}$, $TB_{XY}(\Omega)|_{\Omega=\{\emptyset\}} \geq \overline{TB}_{XY}$ and $TB_{YZ}(\Omega)|_{\Omega=\{RTA_{XY}\}} < \overline{TB}_{YZ}$. By P3,

$$TB_{YZ}(\Omega)|_{\Omega=\{RTA_{XY}\}} = \max_{\Omega}\{TB_{YZ}(\Omega)\}$$

Hence, Y's dominant strategy is "Accept only RTA_{XY}". Similarly, X is always best off rejecting RTA_{XZ}. Thus, by $TB_{XY}(\Omega)|_{\Omega=\{\emptyset\}} \geq \overline{TB}_{XY}$ we have that $\Omega = \{RTA_{XY}\}$ results.

When, instead, $TB_{XY}(\Omega)|_{\Omega=\{\emptyset\}} < \overline{TB}_{XY}$, $TB_{XZ}(\Omega)|_{\Omega=\{RTA_{XY}\}} < \overline{TB}_{XZ}$ and $TB_{YZ}(\Omega)|_{\Omega=\{\emptyset\}} \geq \overline{TB}_{YZ}$ then X's dominant strategy is "Reject both", and, given X's response, Y accepts RTA_{YZ} (as does Z). Therefore, $\Omega = \{RTA_{YZ}\}$ results.

When $TB_{XY}(\Omega)|_{\Omega=\{RTA_{YZ}\}} \geq \overline{TB}_{XY}$, $TB_{XY}(\Omega)|_{\Omega=\{RTA_{XY}\}} < \overline{TB}_{XY}$ and $TB_{YZ}(\Omega)|_{\Omega=\{RTA_{XZ}\}} \geq \overline{TB}_{YZ}$ then by P3

$$TB_{YZ}(\Omega)|_{\Omega=\{RTA_{YZ}\}} = \min_{\Omega}\{TB_{YZ}(\Omega)\}$$

and Y's dominant strategy is "Accept both". Given this action, X's best response is "Accept only RTA_{XY}", leading to $\Omega = \{RTA_{XY}, RTA_{YZ}\}$. ■

Proof of Proposition 2. By the Proof of Proposition 1, whenever country X accepts an agreement, then so do other countries. Furthermore, by P3, when $W_X|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} > W_X|_{\Omega=\{MTA\}} \geq W_X|_{\Omega=\{\emptyset\}}$ then

$$W_X|_{\Omega=\{RTA_{XY}\}} > W_X|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} > W_X|_{\Omega=\{MTA\}} \geq W_X|_{\Omega=\{\emptyset\}}$$

and, therefore, in the standard Game it is the dominant strategy for X to reject the MTA during multilateral negotiations. But in the No-Regionalism Game X is left with $W_X|_{\Omega=\{MTA\}} \geq W_X|_{\Omega=\{\emptyset\}}$ and accepts the MTA. This proves point 1.

Similarly, by P3

$$W_X|_{\Omega=\{RTA_{XY}\}} > W_X|_{\Omega=\{RTA_{XY},RTA_{YZ}\}} \geq W_X|_{\Omega=\{\emptyset\}} > W_X|_{\Omega=\{MTA\}}$$

holds for the setting in point 2. Hence, in the No-Regionalism Game X rejects the MTA, leading to $\Omega = \{\emptyset\}$. But in the standard Game X rejects the MTA, while playing dominant strategy "Accept RTA_{XY}" during RTA negotiations. Hence, in the standard Game $\Omega = \{\emptyset\}$ does not come about. This proves point 2. ■

Proof of Proposition 3. We first prove the statement in point 1 and, subsequently, the

statement in point 2. Given points 1 and 2, any initial conditions $(\delta,\ TB_{ij}^{t=1}(\Omega^{t=1})\big|_{\Omega^{t=1}=\{\emptyset\}}$, $\lim_{t\to\infty}\left[W_i^t|_{\Omega^t=\{RTA_{ij}\}}-W_i^t|_{\Omega^t=\{MTA\}}\right])$ lead to one of the three equilibrium paths described in the statement of Proposition 3. Therefore, these equilibrium paths are the only ones that exist, and Proposition 3 is proven.

We prove point 1 by showing that when $\lim_{t\to\infty}\left[W_i^t|_{\Omega^t=\{RTA_{ij}\}}-W_i^t|_{\Omega^t=\{MTA\}}\right]\leq 0$ then: a) for $t\to\infty$ world free trade (MTA) always obtains; b) in period 1 the MTA is signed only when imbalances are as in point 1 of Proposition 1. Given a) and b), for small enough first period imbalances the MTA holds from period 1 onwards (Good Path), and if imbalances are larger, then regionalism converges to world free trade over time (Ugly Path).

Consider a). By assumption: $\lim_{t\to\infty}TB_{ij}^t(\Omega^t)=0\forall i,j,\Omega$ and balanced trade obtains for $t\to\infty$. Given P3 and $\lim_{t\to\infty}\left[W_i^t|_{\Omega^t=\{RTA_{ij}\}}-W_i^t|_{\Omega^t=\{MTA\}}\right]\leq 0$ country i has the following welfare ranking over Ω^t:

$$W_i^t\big|_{\Omega^t=\left\{\{MTA\}-\left\{RTA_{j\neq i}\right\}\right\}}>W_i^t\big|_{\Omega^t=\{MTA\}}\geq W_i^t\big|_{\Omega^t=\{RTA_{ij}\}}>W_i^t\big|_{\Omega^t=\left\{RTA_{ij},RTA_{j\neq i}\right\}}>W_i^t\big|_{\Omega^t=\left\{RTA_{j\neq i}\right\}}$$

where $\Omega^t=\{\{MTA\}-\{RTA_{j\neq i}\}\}$ stands for the structure that includes the two RTAs to which country i can be part, but excludes $RTA_{j\neq i}$. Moreover, by P1 and P2

$$W_i^t\big|_{\Omega^t=\{RTA_{ij}\}}>W_i^t\big|_{\Omega^t=\{\emptyset\}}$$

Given this ranking, during RTA negotiations country i accepts any RTA that it had not previously signed, regardless of whether $RTA_{j\neq i}$ is in force. As this holds true for all countries, world free trade always obtains for $t\to\infty$. At the period when $\Omega^t=\{RTA_{XY},RTA_{XZ},RTA_{YZ}\}$ is known to be the outcome of RTA negotiations, the MTA is accepted at the preceding MTA negotiations (by the tie-breaking assumption).

Now, consider b). When first period imbalances satisfy the conditions in point 1 of Proposition 1 then, by the Proof of Proposition 1, accepting the MTA is the best response viewed from instantaneous welfare. But all subsequent trade imbalances are smaller than in period 1, by $TB_{ij}^t(\Omega^t)\big|_{\Omega^t=\Omega^{t-1}}=\rho TB_{ij}^{t-1}(\Omega^{t-1})$. Therefore, by P1 and P2, if MTA is optimal in period 1, it is optimal in every subsequent period too.

When, instead, first period imbalances do not satisfy these conditions, country X is best off rejecting at least RTA_{XZ} (Proof of Proposition 1). Moreover, X loses no future welfare by doing so. By P2, Z is always best off accepting the RTA with X. Hence, whenever by $TB_{ij}^t(\Omega^t)\big|_{\Omega^t=\Omega^{t-1}}=\rho TB_{ij}^{t-1}(\Omega^{t-1})$ a point in time arrives when $W_X^t|_{RTA_{XZ}\in\Omega^t}>W_X^t|_{RTA_{XZ}\notin\Omega^t}$, country X can sign the RTA. Gaining instantaneously, but losing no welfare in the future, country X rejects the MTA in period 1. This completes the proof of point b).

Next, we prove point 2. By P3 and $\lim_{t\to\infty}\left[W_i^t|_{\Omega^t=\{RTA_{ij}\}}-W_i^t|_{\Omega^t=\{MTA\}}\right]>0$ we again

obtain country i's welfare ranking over Ω^t for $t \to \infty$:

$$W_i^t\big|_{\Omega^t=\left\{\{MTA\}-\left\{RTA_{j\neq i}\right\}\right\}} > W_i^t\big|_{\Omega^t=\{RTA_{ij}\}} > W_i^t\big|_{\Omega^t=\{MTA\}} > W_i^t\big|_{\Omega^t=\left\{RTA_{ij},RTA_{j\neq i}\right\}} > W_i^t\big|_{\Omega^t=\left\{RTA_{j\neq i}\right\}}$$

Consider the case that there is only one RTA, $\Omega^t = \{RTA_{ij}\}$. For instance, take $\Omega^t = \{RTA_{XY}\}$. Viewed from instantaneous welfare, "Accept RTA_{XZ}" and "Accept RTA_{YZ}" are the dominant strategies for X and Y, respectively: $W_X^t|_{\Omega^t=\{RTA_{XY},RTA_{XZ}\}} > W_X^t|_{\Omega^t=\{RTA_{XY}\}}$ and $W_X^t|_{\Omega^t=\{RTA_{XY},RTA_{XZ},RTA_{YZ}\}} > W_X^t|_{\Omega^t=\{RTA_{XY},RTA_{YZ}\}}$, and likewise for Y. Thus, accepting all RTAs (and, therefore, MTA at the preceding MTA negotiations) is a one-shot Nash Equilibrium.

However, both X and Y are worse off under $\Omega^t = \{MTA\}$ than under $\Omega^t = \{RTA_{XY}\}$. Therefore, they could play a Tit-for-Tat strategy whereby neither signs the RTA with Z as long as the other does not do so either. If one country signs the RTA with Z, then the next period the other country signs it too. Hence, $\Omega^t = \{RTA_{XY}\}$ is sustainable as long as the instantaneous gains from signing the RTA with Z, given that the other does not, are smaller than the subsequent losses from moving to world free trade:

$$W_i^t\big|_{\Omega^t=\left\{\{MTA\}-\left\{RTA_{j\neq i}\right\}\right\}} - W_i^t\big|_{\Omega^t=\{RTA_{ij}\}} < \sum_{s=1}^{\infty} \delta^s \left(W_i^t\big|_{\Omega^t=\{MTA\}} - W_i^t\big|_{\Omega^t=\{RTA_{ij}\}} \right)$$

Thus, there is some $\overline{\delta} \in (0,1)$ such that for $\delta \geq \overline{\delta}$ the Bad Path is sustainable. Of course, even for $\delta \geq \overline{\delta}$ this is not the only equilibrium. If one or both players do not follow a Tit-for-Tat strategy, the MTA is signed. This proves point 2. ■

Proof of Proposition 4. We first prove the statement that world free trade is the unique long run equilibrium. We do this by showing that the alternative structures are not sustainable forever: 1) Complete non-cooperation, $\Omega^t = \{\emptyset\}$; 2) One RTA, $\Omega^t = \{RTA_{ij}\}$; 3) Two RTAs, $\Omega^t = \{\{MTA\} - \{RTA_{j\neq i}\}\}$.

Consider first $\Omega^t = \{\emptyset\}$. By $\rho < 1$ and infinite periods, for any country pair ij, $\exists \bar{t}$: $TB_{ij}^{\bar{t}}\left(\Omega^{\bar{t}}\right) = 0$. Then, by P1 and P2, $W_i^{\bar{t}}\Big|_{\Omega^{\bar{t}}=\{RTA_{ij}\}} > W_i^{\bar{t}}\Big|_{\Omega^{\bar{t}}=\{\emptyset\}} \wedge W_j^{\bar{t}}\Big|_{\Omega^{\bar{t}}=\{RTA_{ij}\}} > W_j^{\bar{t}}\Big|_{\Omega^{\bar{t}}=\{\emptyset\}}$ and likewise, by $E\left[\varepsilon_{ij}^t\right] = 0$ we have that $E^{\bar{t}}\left[W_i^t|_{\Omega^t=\{RTA_{ij}\}} - W_i^t|_{\Omega^t=\{\emptyset\}}\right] > 0 \forall t > \bar{t}$. Therefore, RTA_{ij} is mutually accepted and $\lim_{t\to\infty} \Omega^t \neq \{\emptyset\}$.

Next, consider $\Omega^t = \{RTA_{ij}\}$. In particular, call i's RTA partner j^{in} and the outsider j^{out}. We need to show that there exist $TB_{ij}^t\left(\Omega^t\right)\big|_{\Omega^t=\left\{RTA_{ij^{in}}\right\}}$ for which i and j^{out} are both willing to accept $RTA_{ij^{out}}$. If such realizations exist then, by infinite periods and shocks every period, they will at some point be realized.

First we show that $\exists\, TB_{ij}^t\left(\Omega^t\right)\big|_{\Omega^t=\left\{RTA_{ij^{in}}\right\}}$ for which country i accepts $RTA_{ij^{out}}$. By P1

and P2, $TB^t_{ij^{out}}(\Omega^t) > 0 \Longrightarrow W^t_i|_{\Omega^t=\left\{RTA_{ijin},RTA_{ijout}\right\}} > W^t_i|_{\Omega^t=\left\{RTA_{ijin}\right\}}$. Moreover, by P1,

$$\frac{\partial}{\partial\, TB^t_{ij^{out}}(\Omega^t)\Big|_{\Omega^t=\left\{RTA_{ijin}\right\}}}\left[W^t_i\big|_{\Omega^t=\left\{RTA_{ijin},RTA_{ijout}\right\}} - W^t_i\big|_{\Omega^t=\left\{RTA_{ijin}\right\}}\right] > 0$$

which implies that at a given point in time, $\bar{t}$

$$\frac{\partial}{\partial\, TB^{\bar{t}}_{ij^{out}}\left(\Omega^{\bar{t}}\right)\Big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin}\right\}}}\left[E^{\bar{t}}\sum_{t=\bar{t}}^{\infty}\delta^{t-\bar{t}}\left[W^t_i\big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin},RTA_{ijout}\right\}} - W^t_i\big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin}\right\}}\right]\right] > 0$$

So that the expected gain from an agreement is monotonically increasing in the trade surplus. Hence, $\exists\, TB^{\bar{t}}_{ij^{out}}\left(\Omega^{\bar{t}}\right)\Big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin}\right\}} : E^{\bar{t}}\sum_{\bar{t}}^{\infty}\delta^{t-\bar{t}}\left[W^t_i|_{\Omega^{\bar{t}}=\left\{RTA_{ijin},RTA_{ijout}\right\}} - W^t_i|_{\Omega^{\bar{t}}=\left\{RTA_{ijin}\right\}}\right] > 0$.

Next, we bring in country j^{out}'s decision. P2 and P3 imply that a RTA always raises the joint welfare of two countries. By P2 there are gains from trade, otherwise country i would reject a RTA for any $TB^t_{ij}(\Omega^t) < 0$. While by P3 those gains do not accrue to the outsider. Thus:

$$\left[W^t_i + W^t_{j^{out}}\right]\big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin},RTA_{ijout}\right\}} - \left[W^t_i + W^t_{j^{out}}\right]\big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin}\right\}} > 0$$

As $RTA_{ij^{out}}$ creates joint value, the $TB^{\bar{t}}_{ij^{out}}\left(\Omega^{\bar{t}}\right)\Big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin}\right\}}$ for which country i is better off accepting must be smaller than the $TB^{\bar{t}}_{ij^{out}}\left(\Omega^{\bar{t}}\right)\Big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin}\right\}}$ for which j^{out} is better off rejecting $RTA_{ij^{out}}$. Therefore, there exist $TB^{\bar{t}}_{ij^{out}}\left(\Omega^{\bar{t}}\right)\Big|_{\Omega^{\bar{t}}=\left\{RTA_{ijin}\right\}}$ for which both i and j^{out} accept $RTA_{ij^{out}}$ and $\lim_{t\to\infty}\Omega^t \neq \left\{RTA_{ijin}\right\}$.

Finally, consider $\Omega^t = \left\{RTA_{ijin}, RTA_{ij^{out}}\right\} = \{\{MTA\} - \{RTA_{j\neq i}\}\}$. By P1 and P2,

$$TB^t_{j\neq i}\left(\Omega^t\right) = 0 \Longrightarrow W^t_{j\neq i}\big|_{\Omega^t=\{MTA\}} > W^t_{j\neq i}\big|_{\Omega^t=\left\{\{MTA\}-\left\{RTA_{j\neq i}\right\}\right\}}$$

Then, by the same argument as for $\Omega^t = \{\emptyset\}$, $\exists\bar{t}$ for which both $j \neq i$ are best off accepting $RTA_{j\neq i}$. Hence, $\lim_{t\to\infty}\Omega^t \neq \{\{MTA\} - \{RTA_{j\neq i}\}\}$. Overall, therefore, from any given Ω^t additional RTAs are signed, and hence: $\lim_{t\to\infty}\Omega^t = \{MTA\}$.

Now, we prove the second sentence in Proposition 4. When first period imbalances satisfy point 1 in Proposition 1 then $\Omega^{t=1} = \{MTA\}$. After all, imbalances are then sufficiently small that all countries gain instantaneously (Proof of Proposition 1), and by $\rho < 1$ and $E\left[\varepsilon^t_{ij}\right] = 0$ in expectation imbalances for $t > 1$ are smaller than at $t = 1$. With instantaneous gain and no expected future loss, the MTA is signed. But since $\exists TB^{t=1}_{ij}(\Omega^{t=1}) : \Omega^{t=1} \neq \{MTA\}$ the expected time till $\Omega^t = \{MTA\}$ is (weakly) increasing in first period imbalances. ■

B Microfoundations

This appendix presents a general model that matches properties P1-P3. To do so the following elements will prove sufficient:

1. There are trade imbalances.

2. There is full specialization in production.

3. Tariffs can affect the terms-of-trade.

4. There are gains from trade.

Before presenting the formal derivation, we explain intuitively how these elements lead to properties P1-P3. Trade imbalances in conjunction with the ability to affect terms-of-trade lead to P1: the larger a country's trade deficit is, the more it loses more from giving up the ability to influence its import prices. When, in addition, there are gains from trade, P2 obtains, because there is a trade-off between the gains from trade and the losses from yielding independent tariff setting. Beyond a threshold trade deficit, losses exceed gains. Finally, full specialization and the ability to affect terms-of trade together bring about P3. The reason is that when two countries sign an agreement to cut bilateral tariffs, they raise the price of each other's export goods. Given that the third country does not produce these goods itself, but imports them, the agreement makes it worse off.

Thus, we present a model that contains these four elements and microfounds our one-shot game. The model has two periods. In the first period there are country-specific income shocks and consumers smooth consumption through international borrowing and lending. When a country borrows from abroad, it purchases more than it sells, and is in trade deficit. Upon repayment of the debt it must generate a trade surplus. This is a standard approach to modelling imbalances (Obstfeld and Rogoff (1996)). Subsequently, in the second period, the trade negotiation game described in section 3 takes place.

B.1 Intertemporal preferences

As is common in the literature on trade imbalances, preferences over the intertemporal allocation of consumption expenditure and preferences over the intratemporal allocation of goods are separately defined (Obstfeld and Rogoff (1996)). Intertemporal utility in country i is given by $U\left(C_i^t\right)$ where C_i^t is country i's total consumption expenditure in period t. We do not need to specify a particular functional form. However, we assume that $\frac{dU\left(C_i^t\right)}{dC_i^t} > 0$ and $\frac{d^2U\left(C_i^t\right)}{d\left(C_i^t\right)^2} < 0$:

intertemporal preferences are concave. This bring about the incentive to smooth consumption over time. Country i's representative consumer maximizes:

$$\max_{C_i^{t=1}, C_i^{t=2}} \left\{ U\left(C_i^{t=1}\right) + \delta U\left(C_i^{t=2}\right) \right\} \tag{A.1}$$

where δ is the discount rate. Income in the periods is given by

$$Y_i^t = \overline{Y} + g\left(C_i^t, C_j^t\right) + \nu_i^t \tag{A.2}$$

where $\overline{Y}$ is an exogenous income component and $g\left(C_i^t, C_j^t\right)$ is an income component that depends on consumption decisions in country i and abroad. It includes producer profits and tariff revenues. Finally, ν_i^t is the country-specific income shock with $\nu_i^{t=1}$ being independently and identically distributed across countries, while $\nu_i^{t=2} = 0$ for all countries.

To smooth income shocks consumers are allowed to buy and sell an internationally traded bond, B, which pays interest rate r. Hence,

$$C_i^{t=1} = Y_i^{t=1} - B_i \tag{A.3}$$

and

$$C_i^{t=2} = Y_i^{t=2} + (1+r) B_i \tag{A.4}$$

The bond is the only internationally traded asset. Financial market clearing requires that r adjusts so as to obtain

$$\sum_i B_i = 0 \tag{A.5}$$

We can now rewrite the consumer's optimization problem to

$$\max_{B_i} \left\{ U\left(\overline{Y} + g\left(C_i^{t=1}, C_j^{t=1}\right) + \nu_i^{t=1} - B_i\right) + \delta U\left(\overline{Y} + g\left(C_i^{t=2}, C_j^{t=2}\right) + (1+r) B_i\right) \right\}$$

where the outcome of trade negotiations in period 2 affects $g\left(C_i^{t=2}, C_j^{t=2}\right)$. In fact, we assume that governments play no role in the first period, and that first period tariffs are exogenously given. Governments only play the negotiation game in period 2. By separating consumer and government decisions over the two periods, much complexity is avoided. In the second period governments take consumers' first period decision as given. And in the first period consumers optimize taking into account the effect that their borrowing and lending has on second period negotiations. The first order condition to the consumer's problem yields:

$$\frac{dU\left(\overline{Y} + g\left(\cdot\right)^{t=1} + \nu_i^{t=1} - B_i\right)/dB_i}{dU\left(\overline{Y} + g\left(\cdot\right)^{t=2} + (1+r) B_i\right)/dB_i} = \frac{\delta\left[(1+r) + \frac{d}{dB_i} g\left(\cdot\right)^{t=2}\right]}{1 - \frac{d}{dB_i} g\left(\cdot\right)^{t=1}} \tag{A.6}$$

where the effect of the consumer's first period decision on the outcome of the negotiation game in the second period is contained in $\frac{d}{dB_i} g\left(\cdot\right)^{t=2}$.

Now, take $\nu_i^{t=1} < \nu_j^{t=1} \forall j$. That is, country i experiences a negative income shock relative to other countries. Then, by $\frac{dU\left(C_i^t\right)}{dC_i^t} > 0$ and $\frac{d^2U\left(C_i^t\right)}{d\left(C_i^t\right)^2} < 0$ we have that

$$\frac{d}{dB_i} U\left(\overline{Y} + g\left(\cdot\right)^{t=1} + \nu_i^{t=1} - B_i\right)\Big|_{B_i=0} < \frac{d}{dB_j} U\left(\overline{Y} + g\left(\cdot\right)^{t=1} + \nu_j^{t=1} - B_j\right)\Big|_{B_j=0} \quad \forall j$$

and, therefore, $B_i < 0$ is optimal. Consumers in country i borrow in period 1 and repay in period 2. Notice that it is the relative size of shocks that determines borrowing and lending ($\nu_i^{t=1} < \nu_j^{t=1}$), not the absolute size ($\nu_i^{t=1} \lessgtr 0$). When, say, $\nu_i^{t=1} = \nu_j^{t=1} \forall j$ with $\nu_i^{t=1} < 0$, no borrowing and lending takes place, since all countries are in an identical position.

As trade negotiations occur in the second period, for $\nu_i^{t=1} < \nu_j^{t=1} \forall j$ country i would be in the position of country Z in our one-shot game. Likewise, when $\nu_i^{t=1} > \nu_j^{t=1} \forall j$ then country i is in the position of country X during trade negotiations.

Therefore, borrowing in the first period implies a weaker negotiation position in the second period (since country Z is the least favoured partner - see section 3). Likewise, first period lending can bestow benefits beyond interest payment. This, of course, is factored into consumers' optimal borrowing and lending (through $\frac{d}{dB_i} g\left(\cdot\right)^{t=2}$) and hence into market clearing interest rates.

B.2 Intratemporal preferences

The intertemporal decision determines how much the domestic consumer spends on consumption each period. Intratemporal allocation within a given period faces the following budget constraint, therefore:

$$\sum_k p_{ik}^t c_{ik}^t = C_i^t \tag{A.7}$$

where c_{ik}^t is the consumption in country i of good k at time t, and p_{ik}^t is that good's price. Intratemporal utility over the consumption of the goods, $u\left(c_{ik}^t\right)$, does not require a particular functional form. We assume, however, that preferences generate downward sloping demand

$$\frac{\partial c_{ik}^t}{\partial p_{ik}^t} < 0 \tag{A.8}$$

In addition, we assume that demand elasticities are the same for different goods, and across different countries:

$$\frac{\partial c_{ik}^t}{\partial p_{ik}^t} = \frac{\partial c}{\partial p} \forall i, k, t \tag{A.9}$$

This will ensure that countries' relative market power over different product markets can be expressed in terms of differences in consumption shares only.

B.3 Production

Producers in each country produce one good only, which is made nowhere else. Thus, each country exports one good and imports two, i.e., there is full specialization. We denote by c_{ij}^t and p_{ij}^t the consumption and price in country i of country j's export good. Likewise, c_{ii}^t and p_{ii}^t are the domestic consumption and domestic price of the locally produced good.

Production is subject to decreasing returns to scale and perfect competition. Denote by Q_i^t the total quantity supplied of country i's export good by local producers. Modelling in reduced form:

$$Q_i^t = h\left(p_{ii}^t\right) \tag{A.10}$$

with $h'\left(p_{ii}^t\right) > 0$, while market clearing implies

$$c_{ii}^t + \sum_j c_{ji}^t = Q_i^t \tag{A.11}$$

B.4 Tariffs

The difference between the local price of an import good and the price in its country of origin, $p_{ij}^t - p_{jj}^t$, comes about through import tariffs. In particular,

$$p_{ij}^t = f\left(p_{jj}^t, \tau_{ij}^t\right) \tag{A.12}$$

where τ_{ij}^t is country i's import tariff on country j's good, and $\frac{\partial p_{ij}^t}{\partial \tau_{ij}^t} > 0$. We can show, furthermore, that part of the tariff is absorbed by a lower world price of the good, $\frac{\partial p_{jj}^t}{\partial \tau_{ij}^t} < 0$. This follows from equations (A.8), (A.10), (A.11) and (A.12):

$$\frac{\partial p_{jj}^t}{\partial \tau_{ij}^t} = \frac{dp_{jj}^t}{dQ_j^t}\frac{\partial Q_j^t}{\partial c_{ij}^t}\frac{\partial c_{ij}^t}{\partial p_{ij}^t}\frac{\partial p_{ij}^t}{\partial \tau_{ij}^t} = + + - + < 0 \tag{A.13}$$

which means that tariffs affect the terms-of-trade. Terms-of-trade can be written as the number of goods country i can import with the proceeds from one of its goods being exported, p_{ii}^t/p_{jj}^t (ex-tariff price of exports over ex-tariff price of imports). Thus, a higher tariff of i on j implies that country j's exports become cheaper, and its terms-of-trade worsen, while those of country i improve.

Finally, we assume that the ability to affect terms-of-trade strengthens in the market power

over a good:

$$\frac{\partial^2 p_{jj}^t}{\partial \tau_{ij}^t \partial \left[c_{ij}^t / Q_j^t\right]} < 0 \tag{A.14}$$

The greater the fraction a country consumes of a foreign good, c_{ij}^t/Q_j^t, the greater the ability to depress its price with import tariffs.

We assume, moreover, that governments make optimal use of this market power when they set tariffs (which only happens when negotiations conclude without an agreement in the second period). Standard optimum tariff theory arguments (Johnson (1953-1954)) then imply that

$$\frac{c_{ij}^t}{Q_j^t} > \frac{c_{ji}^t}{Q_i^t} \Longleftrightarrow \left.\frac{p_{ii}^t}{p_{jj}^t}\right|_{RTA_{ij}\in\Omega^t} < \left.\frac{p_{ii}^t}{p_{jj}^t}\right|_{RTA_{ij}\notin\Omega^t} \tag{A.15}$$

When one country has greater influence over terms-of-trade than another country, it has lower terms-of-trade under bilateral free trade than under non-cooperation.

B.5 Results

Properties P1-P3 are written in terms of welfare. To link statements about terms-of-trade to statements about welfare, we first specify that:

$$\frac{\partial W_i^t}{\partial \left[p_{ii}^t / p_{jj}^t\right]} > 0 \tag{A.16}$$

That is, a country's welfare is increasing in its terms-of-trade.

Moreover, by downward sloping demand and upward sloping supply from equations (A.8) and (A.10), standard deadweight losses from taxation imply that

$$\left[W_i^t + W_j^t\right]\Big|_{RTA_{ij}\in\Omega^t} > \left[W_i^t + W_j^t\right]\Big|_{RTA_{ij}\notin\Omega^t} \tag{A.17}$$

and tariff elimination raises joint welfare.

To match P1-P3, welfare statements, in turn, need to be related to the trade balance:

$$TB_{ij}^t = p_{ii}^t c_{ji}^t - p_{jj}^t c_{ij}^t \tag{A.18}$$

where $TB_{ij}^t < 0 \Longleftrightarrow C_i^t > C_j^t$. And, by equation (A.9), $C_i^t > C_j^t \Longleftrightarrow \frac{c_{ij}^t}{Q_j^t} > \frac{c_{ji}^t}{Q_i^t}$. Hence, by equation (A.15), deficit country i suffers a terms-of-trade loss from signing an agreement. It follows that

$$\frac{\partial}{\partial TB_{ij}^t}\left[\left.\frac{p_{ii}^t}{p_{jj}^t}\right|_{RTA_{ij}\in\Omega^t} - \left.\frac{p_{ii}^t}{p_{jj}^t}\right|_{RTA_{ij}\notin\Omega^t}\right] < 0 \tag{A.19}$$

That is, the size of the loss increases in the bilateral trade deficit, from which:

$$\frac{\partial}{\partial TB_{ij}^t}\left[W_i^t\big|_{RTA_{ij}\in\Omega^t} - W_i^t\big|_{RTA_{ij}\notin\Omega^t}\right] > 0 \tag{A.20}$$

which is P1. Thus, a country faces a trade-off between the efficiency gains from an agreement (A.17) and losing its ability to affect terms-of-trade (A.20). In particular, by equation (A.17):

$$\left[W_i^t\big|_{RTA_{ij}\in\Omega^t} - W_i^t\big|_{RTA_{ij}\notin\Omega^t}\right]\Big|_{TB_{ij}^t=0} > 0 \tag{A.21}$$

so that by equation (A.20) there exists some $\overline{TB}_{ij}^t < 0$ at which

$$\left[W_i^t\big|_{RTA_{ij}\in\Omega^t} - W_i^t\big|_{RTA_{ij}\notin\Omega^t}\right]\Big|_{TB_{ij}^t=\overline{TB}_{ij}^t} = 0 \tag{A.22}$$

This is P2. Finally, $RTA_{j\neq i}$ implies that countries $j \neq i$ cut tariffs on each other's goods, which raises mutual consumption of each other's goods and p_{jj}^t increases for both $j \neq i$. Therefore, $\frac{p_{ii}^t}{p_{jj}^t}$ decreases, so that

$$W_i^t\big|_{RTA_{j\neq i}\in\Omega^t} - W_i^t\big|_{RTA_{j\neq i}\notin\Omega^t} < 0 \tag{A.23}$$

and

$$TB_{ij}\big|_{RTA_{j\neq i}\in\Omega^t} - TB_{ij}\big|_{RTA_{j\neq i}\notin\Omega^t} < 0 \tag{A.24}$$

which matches P3.

References

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

[2] Anderson, J.E., and E. van Wincoop (2001) "Borders, Trade and Welfare" NBER Working Paper No. 8515.

[3] Bagwell, K., and R.W. Staiger (1990) "A Theory of Managed Trade" *American Economic Review* 80, 779-795.

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

[5] Bagwell, K., and R.W. Staiger (1997b) "Multilateral Cooperation During the Formation of Free Trade Areas" *International Economic Review* 38, 291-319.

[6] Bagwell, K., and R.W. Staiger (1999) "An Economic Theory of GATT" *American Economic Review* 89, 215-248.

[7] Bagwell, K., and R.W. Staiger (2006) "What Do Trade Negotiators Negotiate About: Evidence from the World Trade Organization" NBER Working Paper Series No. 12727.

[8] Baldwin, R.E., (1995) "A Domino Theory of Regionalism", in R.E. Baldwin, P. Haaparanta and J. Kiander (eds.), *Expanding European Regionalism: The EU's New Members*, Cambridge University Press, Cambridge, MA.

[9] Baldwin, R.E., (1997) "The Causes of Regionalism" *The World Economy*, 865-888.

[10] Baldwin, R.E., (2005) "Stepping Stones or Building Blocs? Regional and Multilateral Integration" in J. McKay, M. Armengol and G. Pineau (eds.), *Regional Integration in a Global Framework*, European Central Bank publication, Frankfurt.

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

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

[13] Bond, E.W., C. Syropoulos, and L.A. Winters (2001) "Deepening of Regional Integration and Multilateral Trade Agreements" *Journal of International Economics* 53, 335-362.

[14] Broda, C., N. Limao, and D.E. Weinstein (2006) "Optimal Tariffs: the Evidence", NBER Working Paper Series No. 12033.

[15] Burbidge, J.B., J.A. DePater, G.M. Myers and A. Sengupta (1997) "A Coalition Formation Approach to Equilibrium Federations and Trading Blocs" *American Economic Review* 87, 940-956.

[16] Chang, W., and L.A. Winters (2002) "How Regional Blocs Affect Excluded Countries: The Price Effects of MERCOSUR" *American Economic Review* 92, 889-904.

[17] Collier, P., (2006) "Why the WTO is Deadlocked: And What Can Be Done About It" *The World Economy*, 1423-1449.

[18] Conconi, P., and C. Perroni (2003) "Self-Enforcing International Agreements and Domestic Policy Credibility" Warwick University Working Paper.

[19] Crawford, J-A., and R.V. Fiorentino (2005) "The Changing Landscape of RTAs" WTO Publication, Geneva.

[20] Dodge, D., (2005) Remarks by David Dodge, Governor of the Bank of Canada, to 'la Conference de Montreal' at Montreal, Quebec, 30 May 2005.

[21] Freund, C., (2000a) "Multilateralism and the Endogenous Formation of Preferential Trade Agreements" *Journal of International Economics* 52, 359-376.

[22] Freund, C., (2000b) "Different Paths to Free Trade: The Gains from Regionalism" *Quarterly Journal of Economics* 115, 1317-1341.

[23] Grossman, G., and E. Helpman (1995) "The Politics of Free-Trade Agreements" *American Economic Review* 85, 667-690.

[24] Holmes, T., (2005) "What Drives Regional Trade Agreements That Work?" HEI Working Paper No: 07/2005.

[25] Horn, H., G. Maggi, and R.W. Staiger (2006) "Trade Agreements as Endogenously Incomplete Contracts" Working Paper.

[26] Johnson, H.S., (1953-1954) "Optimum Tariffs and Retaliation" *Review of Economic Studies* 21, 142-153.

[27] Kennan, J., and R. Riezman (1988) "Do Big Countries Win Tariff Wars?" *International Economic Review* 29, 81-85.

[28] Krishna, P., (1998) "Regionalism and Multilateralism: A Political Economy Approach" *Quarterly Journal of Economics* 113, 227-251.

[29] Krugman, P., (1993) "Regionalism versus Multilateralism: Analytic Notes" in J. de Melo and A. Panagariya (eds.), *New Dimensions in Regional Integration*, Cambridge University Press, Cambridge, MA.

[30] Levy, P.I., (1997) "A Political-Economic Analysis of Free Trade Agreements" *American Economic Review* 87, 506-19.

[31] Ludema, R., (1991) "International Trade Bargaining and the Most Favoured Nation Clause" *Economics and Politics* 3, 1-20.

[32] Lűtkepohl, H., (2005) *New Introduction to Multiple Time Series Analysis*, Springer-Verlag, Berlin.

[33] Lűtkepohl, H. and H.-E. Riemers (1992) "Granger-Causality in Cointegrated VAR Processes: The Case of the Term Structure" *Economics Letters* 40, 263-268.

[34] McLaren, J., (2002) "A Theory of Insidious Regionalism" *Quarterly Journal of Economics* 117, 571-608.

[35] Magee, C.S., (2003) "Endogenous Preferential Trade Agreements: An Empirical Analysis" *Contributions to Economic Analysis and Policy (B.E. Journals)* 2, 1-17.

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

[37] Maggi, G., and A. Rodriguez-Clare (2006) "A Political-Economy Theory of Trade Agreements" Working Paper.

[38] Obstfeld, M., and K. Rogoff (1996) *Foundations of International Macroeconomics*, MIT Press, Cambridge, MA.

[39] Olarreaga, M., I. Soloaga and L.A. Winters (1999) "What's Behind MERCOSUR's Common External Tariff?" World Bank Policy Research Working Paper No. 2231.

[40] Ornelas, E., (2005) "Rent Destruction and the Political Viability of Free Trade Agreements" *Quarterly Journal of Economics* 120, 1475-1506.

[41] Panagariya, A., (2000) "Preferential Trade Liberalization: The Traditional Theory and New Developments" *Journal of Economic Literature*, 38, 287-331.

[42] De Rato, R., (2005) "It is Not Just Up to Washington to Correct Global Imbalances" *European Affairs* 6.

[43] Riezman, R., (1991) "Dynamic Tariffs with Asymmetric Information" *Journal of International Economics*, 30, 267-283.

[44] Saggi, K., and H.M. Yildiz (2006) "Bilateral Trade Agreements and the Feasibility of Multilateral Free Trade" Working Paper.

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

[46] Whalley, J., (2006) "Recent Regional Agreements: Why So Many, Why So Much Variance in Form, Why Coming So Fast, and Where Are They Headed?" CESifo Working Paper No. 1790.

[47] Winters, L.A., (1996) "Regionalism Versus Multilateralism" Policy Research Working Paper 1687, World Bank.

[48] Zissimos, B., (2007 forthcoming) "The GATT and Gradualism" *Journal of International Economics*.