THE GATT AND GRADUALISM

by

Ben Zissimos

Working Paper No. 06-W19

August 2006

DEPARTMENT OF ECONOMICS
VANDERBILT UNIVERSITY
NASHVILLE, TN 37235

www.vanderbilt.edu/econ

The GATT and gradualism[1]

Ben Zissimos, *Vanderbilt University*[2]

First version: April 2002

Revised August 2006

Abstract: This paper shows how the institutional rules imposed on its signatories by the GATT created a strategic incentive for countries to liberalize gradually. Trade liberalization must be gradual, and free trade can never be achieved, if punishment for deviation from an agreement is limited to a 'withdrawal of equivalent concessions' and if initial deviation from an agreement is also limited. The paper shows how (sufficiently patient) countries have an incentive to deviate in a limited way when operating under GATT dispute settlement procedures.

Keywords: free trade; gradual trade liberalization; strategic interactions; trade agreement; welfare.

JEL Classification Numbers: F02, F13, F15, C73.

---

[1]An earlier version of this paper was circulated under the title 'Trade agreements with limited punishments.' I am especially grateful to Ben Lockwood for initiating the project that gave rise to this present paper. I would like to thank two anonymous referees, Rick Bond, Andrew Daughety, Jayasri Dutta, Jennifer Reinganum, Robert Staiger, Quan Wen, John Whalley, Kar-yiu Wong and Isleide Zissimos for helpful comments and conversations about this paper. I am also grateful for comments from seminar participants at the Midwest International Economics Meeting, University of Wisconsin, The WTO and World Trade II Meeting, University of Washington, and the North American Summer Meeting of the Econometric Society, Brown University. Support from the ESRC's Evolving Macroeconomy Programme, The Warwick Centre for Public Economics, and The Centre for the Study of Globalisation and Regionalisation, University of Warwick is gratefully acknowledged.

[2]Dept. of Economics, Vanderbilt University, Nashville, TN 37235.
Tel: ++1 615 322 3339.
Fax: ++1 615 343 8495.
E-mail: ben.zissimos@vanderbilt.edu

1. Introduction

The experience of trade liberalization in the period since World War II has presented economists with two puzzles. First, even in developed countries, free trade has remained stubbornly elusive, with average trade-weighted tariffs remaining at low but still positive levels.[3] Second, tariffs have been cut only gradually under the General Agreement of Tariffs and Trade (GATT). Since the GATT was drawn up after the war, tariffs have fallen from a trade weighted average of 50 percent to around 5 percent today. Neither of these two facts sits well with the textbook view that sees a trade agreement as a simple repeated Prisoners' Dilemma: That is, as a situation where it is individually rational for countries to impose tariffs, but collectively rational to abolish them.

The purpose of this present paper is to propose an explanation for these two puzzles by focusing on the institutional features of the rules imposed on trade liberalization by the GATT. Within this institutional setting, the paper examines the implications for the trade liberalization process of the dispute settlement system under the GATT, particularly a *withdrawal of equivalent concessions* (WEC). In doing so, it argues that the incentives created by these rules were sufficient to motivate the outcomes of gradual liberalization and failure to reach free trade actually observed.

The focus of this paper is on the broad sweep of trade liberalization under the GATT in the post war period, up to the conclusion of the Uruguay Round in 1994. The idea is to assume that two countries have signed the GATT, and then analyze the dynamic equilibrium (liberalization) path that results when a tariff reduction game is played according to GATT rules. The paper undertakes a formal representation of the GATT Articles in question, setting them out in a fully specified (game theoretic, general equilibrium) framework. This includes a formal statement of the GATT dispute settlement procedures that govern deviation from an agreement and the degree of retaliation allowed. Thus, not only is it possible to characterize the liberalization process when liberalization takes place according to the Articles. It is also possible to show that, having signed up to the GATT,

[3] It could be argued that there is no puzzle in the failure to reach free trade. For example, in a world where tariffs are used for political/redistributive purposes, current positive tariff levels could be efficient. However, in practice there appears to be a consensus that efficiency has not been reached; that mutual gains from trade are still available from further multilateral trade liberalization. To keep things simple, free trade will be used as a metaphor for this 'yet to be obtained' efficient level of international trade.

countries could do no better than liberalize according to the Articles.[4] This property of the theoretical framework accords well with the historical period from the GATT's inception in 1947 to the conclusion of the Uruguay Round. Over that time, the GATT Articles were adhered to quite closely. For example, violations of tariff bindings were not often observed; see Chapter 2 of Whalley and Hamilton (1996) for further details.[5]

It is well recognized that there is a basic tension in the dispute settlement procedures of the GATT, which the model of this present paper is designed to capture. The tension seems to have been created by two (not necessarily consistent) purposes that the dispute settlement procedures were designed to serve. The first purpose was to secure compliance with GATT obligations through the threat of punitive sanctions. But the second purpose was that these procedures should be used primarily to maintain the balance of concessions and to avoid the use of punitive sanctions. During the original drafting of the GATT's dispute settlement procedures, statements were made about the need for different kinds of responses to disputes depending on the severity, with "more rigorous retorsion" reserved only for circumstances where the offending action was "abusive" (see Jackson 1969, p. 169 and Staiger 1995a, p. 1501).

The formalization of the tension in the dispute settlement procedures presented in this paper is as follows. Suppose that a deviating country fails to implement some agreed market access measure in a way that is deemed to be non-abusive. Under WEC, contracting parties to the agreement do no more than to withdraw market access concessions equivalent to those that the deviator failed to implement. More rigorous retorsion may be triggered in one of two ways. First, given an initial (non-abusive) deviation from an agreement, if a punisher punishes more aggressively than is allowed by WEC then a standard trigger-strategy punishment is invoked.[6] Second, a deviator cannot expect to be protected

[4]In formal terms, we will show that in all cases the efficient equilibrium path is subgame perfect, and that given a deviation the punishment path is also subgame perfect.

[5]In 1994, as part of the conclusion to the Uruguay Round, signatories to the GATT formed the World Trade Organization (WTO). To some extent the analysis of the present paper is relevant for the period since 1994 too, because the GATT Articles were adopted in the Charter of the WTO (GATT 1994). But since the WTO's inception, we appear to be observing a change in the operation of the regime, with a number of instances where rules have been broken. The reasons for this change present an important agenda for future research, but will not be taken up here in this present paper.

[6]By breaking WEC, Bagwell and Staiger (2002 Chapter 6, p. 99) argue that a government 'ploughs over the final backstop of a GATT ruling' which will lead to a break-down of cooperation, formally equated to Nash reversion or 'trigger strategies.'

by WEC if it undertakes an "abusive" deviation, and in this case Nash reversion is also triggered.[7] This penalty structure is modelled in the context of a dynamic game.

There is no formal definition of "abusive" in the GATT. So to complete the formalization of the present institutional setting, an interpretive assumption about what constitutes an "abusive" deviation will be introduced. A reasonable interpretation would be that an abusive deviation is one which entails a "sufficiently deep" breakage of commitments. Specifically, this could be interpreted as breaking commitments that had actually been honored for some time, as opposed to further commitments that had yet to be attained but were scheduled to be honored at some future date. In the model we shall adopt a particularly tractable form of this assumption, which plays on the simple distinction between commitments that have been honored for at least one period versus commitments that have yet to be implemented.

The first main result of this paper (Proposition 4) incorporates these restrictions on punishers and deviators and, in doing so, characterizes gradualism. The result shows that if punishments are limited to WEC, and if the initial deviation by any country is also limited, then any efficient self enforcing path of trade liberalization is gradual. Because punishment is limited, current tariff cuts can only be made self-enforcing by the promise of future tariff reductions. But if initial deviation is sufficiently limited as well, then it is always possible to promise liberalization over a number of future periods that would more than compensate. So on the efficient equilibrium path, trade liberalization must take place over a number of periods.

The second main result (Proposition 5) is that free trade certainly cannot be reached when trade liberalization is gradual, no matter how little countries discount the future. This result contrasts markedly with conventional insights from the theory of repeated games, which indicate that free trade can be achieved given sufficiently little discounting.

[7] This paper studies 'off-equilibrium-path' retaliation, where there can be no return to the equilibrium tariff reduction path once a deviation has occurred. Bagwell and Staiger point out that aspects of Articles XIX and XXIII sanction both on-equilibrium-path and off-equilibrium-path retaliation, depending on the circumstance (see Bagwell and Staiger 2002, especially p. 97). In contrast to the present paper, Bagwell and Staiger (1990) study on-equilibrium path retaliation. They look at a situation where efficiency cannot be enforced because governments are not sufficiently patient, and then show how 'escape clauses' under Article XIX of the GATT can rebalance an agreement, relaxing the incentive to deviate. See Appendix A.2 of the present paper for further details. Thus, they present an alternative formalization of the basic tension in the GATT dispute settlement procedures discussed above, and their focus is not on gradualism.

The intuition behind this second main result is simple. From the first result, if deviation is limited under the GATT then it is always possible to promise liberalization over a number of future periods that would more than compensate. But more than this, when punishment is limited, for any agreed tariff rate, it is actually *necessary* to promise further liberalization in future periods on the equilibrium path to compensate for potential gains to deviation today. At free trade, such promises of future liberalization are of course impossible.

This second result is related to the so called bicycle theory of trade negotiations discussed by Bhagwati (1988) and formalized by Staiger (1995b). According to the bicycle theory, if a current round of trade negotiations were to fail then there would be a collapse back to higher levels of protectionism. According to the second result of this present paper, if the prospect of future trade liberalization fails then currently scheduled liberalization must fail as well. In both cases, the prospect of future liberalization is required to make current tariff levels self-enforcing. But a collapse backwards does not happen in the framework of this present paper because the scope for deviation is limited.

The paper builds on a substantial literature going back to Johnson's (1953-54) characterization of a tariff war.[8] Later contributions explain trade liberalization in a standard repeated game setting, where tariff cuts from their static Nash equilibrium values are explained as the outcome of self-enforcing trigger strategies (Dixit 1987).[9] As remarked above, using a trigger strategy has two limitations. It cannot explain gradualism and, moreover, free trade is always a self-enforcing outcome with sufficiently little discounting.

Prominent in the literature on which this paper builds is Bagwell and Staiger's (1999) (game theoretic, general equilibrium) framework for analysis of the GATT; elements of their framework are developed in their other papers of 1990, 1996 and 2001.[10] In terms of

[8]Hamilton and Whalley (1983) broaden considerably the basis on which tariff wars can be examined by showing how they can be studied using numerical simulations. Lockwood and Wong (2000) compare trade wars with specific and ad valorem tariffs, showing the outcomes to be different under the respective instruments. Syropoulos (2002) examines the effect of country size, showing that if one trade partner is larger than another by a significantly large ratio, then it will prefer a trade war to free trade.

[9]Among many others, some contributions to the literature on trade agreements that use the threat of retaliation as threat points in cooperative or non-cooperative models include Mayer (1981), Bagwell and Staiger (1990) and McLaren (1997).

[10]For a synthesis, which introduces, develops and extends their approach to a comprehensive treatment of the GATT/WTO as an economic institution, see Bagwell and Staiger (2002).

the theoretical framework developed in the present paper, the most important foundations are laid by Lockwood and Thomas (2002), who study the effect of *complete* irreversibility of strategic actions in an abstract voluntary contributions game. They show that complete irreversibility is sufficient to motivate a gradual increase in contributions. In this present paper we will extend the framework of Lockwood and Thomas from an abstract game with complete irreversibility to a tariff game in which WEC may be thought of as imposing *partial* irreversibility.[11]

Other recent literature has offered several explanations as to why self-enforcing tariff agreements are gradual. The general idea is that, initially, full liberalization cannot be self-enforcing, because the benefits of deviating from free trade are too great to be dominated by any credible punishment. But if there is partial liberalization, structural economic change within the domestic economy reduces the benefits of deviation from further trade liberalization (and/or raises the costs of punishment to the deviator). The individual papers differ in their description of the structural change induced by partial liberalization. Staiger (1995b) endows workers in the import competing sector with specific skills, making them more productive there than elsewhere in the economy. When they move out of this sector they lose their skills with some probability, relaxing the constraint on further trade liberalization. In Devereux (1997), there is dynamic learning-by-doing in the export sector. In Furusawa and Lai (1999), linear adjustment costs are incurred when labor moves between sectors.[12] Bond and Park (2002) show that gradualism arises as a result of asymmetry in country size. In Chisik (2003), increasing sunk costs of investment in the expanding export sector raise the costs of deviation, and increased specialization gradually lowers the lowest obtainable self-enforcing tariff. Conconi and Perroni (2004) show that trade liberalization must be gradual in a self-enforcing trade agreement between a large country and a small country, where the only motive for the trade agreement is a domestic commitment issue that affects the small country. In Maggi and Rodriguez-Clare (2005),

[11] In a two player prisoners' dilemma with continuous actions, under *complete irreversibility* once players have achieved a given level of cooperation neither can reverse their action in order to punish the other. Under *partial irreversibility*, some reversal of actions is possible. WEC entails partial irreversibility in the sense that a punisher is allowed to go back on the level of cooperation achieved, but only to the level of the deviant and no more. Note that an important feature of Lockwood and Thomas (2002) is that their gradualism results do not depend on the discount rate, while those of the present paper do.

[12] Furusawa and Lai have an appendix where they show that with strictly convex adjustment costs, a social planner would choose gradual tariff reduction.

gradualism results from the interaction between frictions in capital mobility and lobbying by capital owners. All of these papers focus on elements of the domestic economy to motivate gradualism, as opposed to elements of the international trading system that we study here.

The paper proceeds as follows. The next section sets up the basic analytical framework, defining formally the tariff reduction game and the punishments for deviation allowed under the GATT, focusing primarily on WEC. Section 3 then studies the effect of WEC on punishment. To do so, Section 3 introduces a standard repeated game framework, the only modification of which is to stipulate that WEC may be adopted under punishment instead of the usual trigger strategies. Here it is shown that WEC represents a subgame perfect equilibrium punishment strategy given sufficiently little discounting. Section 4 then argues that the actions not just of punishers but of initial deviators are limited since (sufficiently patient) countries do better under deviation if they are protected by WEC. Limits on the actions of deviators are introduced (and limits on punishers in the form of WEC are retained) and equilibrium is characterized in the two main results. Conclusions are drawn in Section 5. It is worth emphasizing that the model presented in this paper does not provide a reason for why WEC was included in the GATT. The concluding section provides a brief discussion of this issue and directions for future research that might make progress in understanding its inclusion.

2. Optimal tariffs, trade agreements and punishments under GATT

2.1. Tariffs and welfare

We will work with a standard model of international trade in which two countries produce and consume two final goods.[13] Country i is assumed to have a comparative advantage in the production of commodity i, $i \in \{1, 2\}$. Countries are symmetrical in all other respects. In particular, preferences and technologies are identical subject to a re-ordering of the goods. Both countries are large enough to affect the terms of trade. Import tariffs are the only form of trade restriction allowed; τ_t^i represents the tariff set by i on imports

[13] As Bagwell and Staiger (1999) point out, a two country model is sufficient to carry out an examination of the limits to reciprocal trade liberalization which is the subject here. A model incorporating more countries is only required when issues of non-discrimination against third parties are under discussion.

from j in period t. Tariff revenues are redistributed to citizens in lump-sum. Preferences of the representative consumer in each country over both goods are given by a strictly quasiconcave utility function. The two countries' production possibility loci are strictly concave to the origin over the two goods.[14]

Within a period, the order of events is as follows. First, each country i, observing the tariffs set by the other country (and their own) up to the previous period, simultaneously chooses an import tariff. Then, given world prices of the two goods and tariffs, perfect competition in production takes place. Next, the representative consumer in each country chooses consumption to maximize utility subject to budget constraints. This yields the usual indirect utility function and excess demands. Then, conditional on tariffs, markets clear and world prices for the goods are determined.[15] World prices will of course depend on tariffs, as will tariff revenues. We assume that equilibrium prices are unique.

So, we can write equilibrium welfare of any country i as a function of tariffs only. Let τ be the tariff that country i levies on its imports, and let τ' be the tariff that its exports face upon entry to the foreign market; tariffs are ad valorem. Let $(\tau, \tau') \in \mathbb{R}_+^2$; tariffs cannot be negative.[16] In equilibrium, the indirect utility function can be written $w(\tau, \tau')$. Now we can define a *Nash equilibrium in tariffs* in the usual way as a $\hat{\tau}$ such that $w(\hat{\tau}, \hat{\tau}) \geq w(\tau, \hat{\tau})$ all $\tau \in \mathbb{R}_+$, $i \in \{1, 2\}$. We will focus on symmetric Nash equilibria with trade.[17] Such equilibria exist and are unique for the special cases that we consider below, due to the symmetry of the model. The game described here is obviously static. So, $\hat{\tau}$ may be referred to as the *static Nash equilibrium tariff rate.*

[14] This general specification also encompasses 'endowment models of international trade.' All the general analysis of the paper is worked out for an endowment model example in Appendix A.3.

[15] Prices are determined only up to a scalar, and so a choice of numeraire must be made.

[16] This framework could easily be extended to allow for the possibility of import subsidies as well. However, we will assume that tariffs cannot be negative in order to focus attention on the process of tariff reduction, which is assumed to start with (symmetrical) positive tariffs. Also, in general trade subsidies introduce a number of separate issues which we want to leave aside here; see Jackson (1989). In order to check that the assumption of non-negative tariffs is not restrictive, in the proof of Proposition 5 the assumption is lifted allowing a tariff reduction path in which tariffs are negative to be considered as a candidate for equilibrium. It is then verified that such a path cannot be an efficient equilibrium tariff reduction path.

[17] As discussed by Dixit (1987), there is at least one Nash equilibrium with positive trade flows and a continuum of Nash equilibria at which trade is eliminated.

We assume three properties of $w(\tau, \tau')$:[18]

A1. $w(\tau, \tau')$ is continuous, and strictly concave in τ for all $(\tau, \tau') \in \mathbb{R}^2_+$, with $w_1(0, \tau') > 0$, $\lim_{\tau \to \infty} w_1(\tau, \tau') < 0$, and $w_2(\tau, \tau') < 0$ if $\tau' > 0$.

A2. $w_1(\tau, \tau) + w_2(\tau, \tau) < 0$ for all $\tau > 0$ and $w_1(0,0) + w_2(0,0) = 0$.

A3. $w_{12}(\tau, \tau') < 0$, all $(\tau, \tau') \in \mathbb{R}^2_+$.

A1 and A3, taken together, imply that the static tariff game has a Prisoners' Dilemma structure and has a unique Nash equilibrium (with trade). Under A3, tariffs are strategic substitutes; as one country increases its tariff, the payoff to the other country from increasing its tariff is reduced. A2 says that any equal reduction in all tariffs makes a country better off, and that free trade is optimal.

2.2. Trade agreements

We are interested in how fast countries can reduce tariffs from the static Nash equilibrium, and also whether they can ever reach free trade i.e. $\tau_t^1 = \tau_t^2 = 0$, if the tariff reduction plan must be *self-enforcing* i.e. the outcome of a subgame-perfect equilibrium. Payoffs over the infinite horizon are discounted by a common discount factor δ, $0 < \delta < 1$ i.e.

$$(1-\delta) \sum_{t=0}^{\infty} \delta^t w(\tau_t^i, \tau_t^j). \tag{2.1}$$

A *tariff history* at time t is defined as a complete description of all past tariffs in both countries; $h_t = \{(\tau_1^1, ..., \tau_{t-1}^1, \tau_1^2, ..., \tau_{t-1}^2)\}$. Both countries can observe tariff histories. A *tariff strategy* for country $i \in \{1, 2\}$ is defined as a choice of tariffs τ_t^i in periods $t = 1, 2...$ conditional on every possible tariff history. A *tariff path* of the game is a sequence $\{(\tau_t^1, \tau_t^2)\}_{t=1}^{\infty}$ that is generated by the tariff strategies of both countries.

Given the symmetry of the model, we restrict attention to *symmetric* equilibrium[19] tariff paths where $\tau_t^i = \tau_t$, $t = 1, 2, \ldots$, i.e. where both countries choose the same tariff

[18] We shall write $\partial w(\tau, \tau')/\partial \tau$ as $w_1(\tau, \tau')$, $\partial w(\tau, \tau')/\partial \tau'$ as $w_2(\tau, \tau')$, $\partial^2 w(\tau, \tau')/\partial \tau^2$ as $w_{11}(\tau, \tau')$, $\partial^2 w(\tau, \tau')/\partial \tau \partial \tau'$ as $w_{12}(\tau, \tau')$, and so on.

[19] In the sequel, it is understood that "equilibrium" refers to subgame-perfect Nash equilibrium.

in every time period, and we denote such paths by the sequence $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$, where $\widetilde{\tau}_t$ is the tariff "agreed" for period t and the sequence is initialized at $\widetilde{\tau}_0 = \hat{\tau}$.

2.3. Punishments under the GATT

In a standard repeated tariff game, the punishment that i levies on j for deviating from $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$ is to raise its tariff to the static Nash equilibrium tariff rate $\hat{\tau}$ and maintain it there (which we also call a *trigger strategy*).[20]

In practice, GATT signatories were bound by Article XXIII to adopt a *withdrawal of equivalent concessions* (WEC). Under the WEC rule country i, upon observing that j has deviated at time t, withdraws precisely the equivalent concessions to market access at time $t+1$. That is, if the deviator j has set $\tau_t^j = \tau' > \widetilde{\tau}_t$, then in the next period instead of retaliating by setting $\hat{\tau}$ the other party is allowed to withdraw the concession made, implementing $\tau_{t+1}^i = \tau' = \tau_t^j$ as well.[21] The aim is to analyze equilibrium paths that result when countries are bound by WEC as opposed to standard trigger strategies.

In addition, we must specify what would happen if a country did not adhere to WEC when it punished another country for deviation. Under Article XXIII contracting parties were given the power, 'in appropriately serious cases, to authorize a contracting party or parties to suspend GATT obligations to other contracting parties' (Jackson 1989, p. 94). In the present stylized framework, we will say that to *break WEC*, by setting $\tau_{t+1}^i > \tau' = \tau_t^j > \widetilde{\tau}_t$, results in an *indefinite suspension of GATT obligations* among both parties; there is an indefinite return to trigger strategies $\hat{\tau}$.

[20] In principle, other (more severe) punishments exist such as the autarky Nash equilibrium and Abreu optimal punishment (which are equivalent in this model). These will not be considered in the present setting because they are so costly to carry out and so are vulnerable to renegotiation. (Renegotiation proofness will be discussed further in due course.)

[21] Note that this formalization of WEC depends on the symmetric environment in which we are working. This formalization would not be correct if countries could be asymmetric, since an equal change in tariffs would not then lead to an equivalent change in market access.

3. A withdrawal of equivalent concessions

The main aim of this section is to study how WEC is implemented. We can use the framework set out above to do this. The only difference between the above framework and a standard repeated game is that here countries can punish by WEC instead of using the usual trigger strategies. In this framework, a punisher can also provoke a return to trigger strategies by bringing about a suspension of GATT obligations. However, we will show that sufficiently patient countries prefer to use WEC. That is to say, we will show that WEC is a subgame perfect punishment strategy given sufficiently little discounting of the future.

The ultimate aim of seeing how WEC is implemented is to work out the equilibrium tariff reduction path $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$ that it can support. In the next section (Section 4) we will characterize the equilibrium tariff reduction path under the *additional assumption* that initial deviations are also limited. (Recall from the Introduction the view that deviations were limited by the institutional feature that a deviator could not expect to be protected by WEC if its deviation was deemed to be "abusive.") But in this section (Section 3), in order to isolate the effects of WEC, no restrictions will be placed on initial deviations. Thus we will focus on *optimal deviations* and consider only the off-equilibrium tariff path.[22]

Let us begin by characterizing the optimal deviation from a symmetric equilibrium path $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$ for any country i, given that it rationally anticipates that it will be punished by the WEC rule. Let i's optimal deviation at t from the reference path $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$ be denoted z_t. The payoff from a deviation to z_t at t, denoted $\Delta(z_t, \{\widetilde{\tau}_s\}_{s=t}^{\infty})$, is:

$$\Delta(z_t, \{\widetilde{\tau}_s\}_{s=t}^{\infty}) = \begin{cases} (1-\delta)w(z_t, \tilde{\tau}_t) + (1-\delta)\sum_{s=t+1}^{\infty}\delta^{s-t}w(\tilde{\tau}_s, \tilde{\tau}_s) & \text{if } \widetilde{\tau}_t > z_t \\ (1-\delta)w(z_t, \tilde{\tau}_t) + \delta w(z_t, z_t) & \text{if } \hat{\tau} > z_t > \widetilde{\tau}_t \\ (1-\delta)w(z_t, \tilde{\tau}_t) + \delta w(\hat{\tau}, \hat{\tau}) & \text{if } z_t \geq \hat{\tau} > \widetilde{\tau}_t \end{cases} \tag{3.1}$$

The first line of (3.1) shows the payoff that a country obtains from a deviation by setting a tariff *below* the agreed level; $\widetilde{\tau}_t > z_t$. After such a deviation, there is simply a return

[22] For completeness, at the end of this section I characterize informally the (on-) equilibrium tariff reduction path when no restrictions are placed on initial deviations.

to the agreed tariff path in all future periods. This happens because WEC applies only to deviation by setting a tariff *above* the agreed rate $\widetilde{\tau}_t$. Under WEC, no punishment is imposed if a country deviates by cutting tariffs below the agreed level.

The second line shows the payoff obtained from a deviation above the agreed tariff rate but below the static Nash equilibrium tariff rate; $\hat{\tau} > z_t > \widetilde{\tau}_t$. In the period following the deviation, the punisher also implements z_t, and both countries implement z_t in all future periods in accordance with WEC.[23]

The third line shows the payoff obtained from a deviation such that $z_t > \hat{\tau} > \widetilde{\tau}_t$. In this case, the WEC rule would allow the punisher to implement z_t as well. But by standard arguments, neither country can credibly commit to $z_t > \hat{\tau}$ in future periods and both countries implement $\hat{\tau}$ from $t+1$ onwards. The outcome is exactly the same as in a standard case of trigger strategies.[24]

We are interested in the *optimal deviation* z_t, i.e. the choice of z_t that maximizes $\Delta(z_t, \{\widetilde{\tau}_s\}_{s=t}^{\infty})$ given the reference path. The largest possible gain from deviation is the supremum of $\Delta(z_t, \{\widetilde{\tau}_s\}_{s=t}^{\infty})$ across all values of $z_t \neq \widetilde{\tau}_t$, denoted by $\overline{\Delta}(\{\widetilde{\tau}_s\}_{s=t}^{\infty})$.

Lemma 1. *Assume A1-A2. Then,*

$$\overline{\Delta}(\{\widetilde{\tau}_s\}_{s=t}^{\infty}) =$$

$$\max\{\max_{\infty > z_t \geq \tilde{\tau}_t} [(1-\delta)w(z_t, \tilde{\tau}_t) + \delta w(\min\{z_t, \hat{\tau}\}, \min\{z_t, \hat{\tau}\})],\ (1-\delta)\sum_{s=t}^{\infty} \delta^{s-t} w(\tilde{\tau}_s, \tilde{\tau}_s)\}.$$

[23]We will show in Proposition 1 below that (3.1) generates a punishment path that is subgame perfect. That is, if the initial deviator deviates to $z_t < \hat{\tau}$ and then both countries implement z_t in all future periods then this is a subgame perfect strategy profile. We will also show that, for sufficiently little discounting, WEC is a subgame perfect punishment strategy given the alternative of breaking WEC and provoking an indefinite suspension of GATT obligations (Proposition 2).

[24]We will follow the relatively simplistic approach of assuming that punishments are imposed symmetrically. This assumption has been criticized because all parties have an incentive to 'renegotiate away from' these punishments during the punishment phase. Farrell and Maskin (1989) and van Damme (1989) show that punishments must be asymmetric in order for the punisher to do better under punishment than any feasible alternative. However, following Blume (1994) and McCutcheon (1997) I argue that symmetric punishments can be used to enforce a trade agreement if renegotiation away from punishments sanctioned by the GATT is sufficiently costly. Such renegotiation of the rules is outside the authority of trade negotiators and is more costly than negotiation 'within the rules,' requiring ratification in national legislatures rather than routine rubber stamping by trade negotiators. McCutcheon shows that when renegotiation is more costly than the stipulated trigger strategies then trigger strategies may be used to sustain an agreement. Renegotiation proofness requires finite punishments. Use of infinite punishments is a simplifying but inessential short-cut.

This result says that the best a country can do is either to replicate the payoff on the equilibrium path - the second term in curly brackets - or to deviate by setting tariffs above the agreed level; $z_t \geq \tilde{\tau}_t$. It can never benefit by a unilateral deviation $z_t < \tilde{\tau}_t$.[25]

Having ruled out the possibility of a tariff reduction under deviation, we can now focus on the first term in curly brackets in order to characterize an optimal deviation. Ignoring the inequality constraint for the time being, from the first term in curly brackets define the following functions:

$$\mu(\tilde{\tau}_t) = \arg\max_{z_t} \{(1-\delta)w(z_t, \tilde{\tau}_t) + \delta w(\hat{\tau}, \hat{\tau})\};$$

and

$$\zeta(\tilde{\tau}_t) = \arg\max_{z_t} \{(1-\delta)w(z_t, \tilde{\tau}_t) + \delta w(z_t, z_t)\}. \tag{3.2}$$

These functions are illustrated in Fig. 1. By A3, the functions $\mu(\widetilde{\tau}_t)$ and $\zeta(\tilde{\tau}_t)$ are downward sloping. In particular, note that

$$\zeta'(\tilde{\tau}_t) = \frac{(1-\delta)w_{12}(z_t, \tilde{\tau}_t)}{D}$$

where $D > 0$ from the second-order condition for the choice of z_t in (3.2).

The function $\mu(\tilde{\tau}_t)$ determines the optimal deviation from $\tilde{\tau}_t$ in a standard repeated game, given that the continuation game in the punishment phase has symmetrical tariffs fixed at $\hat{\tau}$.[26] The function $\zeta(\tilde{\tau}_t)$ determines the optimal deviation from $\tilde{\tau}_t$ when punishment is by a withdrawal of equivalent concessions. As Fig. 1 shows, $\zeta(\tilde{\tau}_t)$ lies everywhere below $\mu(\tilde{\tau}_t)$. Under WEC, the size of the deviation determines the size of the punishment. The deviator takes this into account, deviating less aggressively (i.e. by setting a lower tariff) than when the punishment is fixed at $\hat{\tau}$. The less patient countries are, reflected in a lower value of δ, the closer is $\zeta(\tilde{\tau}_t)$ to $\mu(\tilde{\tau}_t)$.[27]

[25] To see why, recall that a withdrawal of equivalent concessions applies only to upward deviations. If a country were to deviate by setting a tariff that were lower than agreed - $z_t < \widetilde{\tau}_t$ - the WEC rule would not require all other countries to follow the deviant downwards. We can therefore ignore the possibility that $z_t < \widetilde{\tau}_t$ because, by A1, a country would make itself worse off by deviating in this way.

[26] The function $\mu(\widetilde{\tau}_t)$ is equivalent to the reaction function of the static game.

[27] Note $\mu(\widetilde{\tau}_t)$ and $\zeta(\widetilde{\tau}_t)$ are not necessarily linear. They are depicted as such in the figures only for clarity. The point of emphasis is that $\mu(\widetilde{\tau}_t)$ lies everywhere above $\zeta(\widetilde{\tau}_t)$. (See Lemma 2 for proof.) Also, to avoid clutter, the figures show the "reaction functions" for only one country.

By symmetry, $\hat{\tau}$ is determined by the intersection of $\mu(\tilde{\tau}_t)$ with the 45-degree line. Define $\overline{\tau}$ to satisfy:

$$\overline{\tau} = \zeta(\overline{\tau}). \tag{3.3}$$

This is a self-enforcing tariff level for $\zeta(\tilde{\tau}_t)$ i.e. at $\overline{\tau}$ the optimal deviation is in fact not to deviate at all. Both $\hat{\tau}$ and $\overline{\tau}$ are illustrated in Fig. 1.

From the functions $\mu(\tilde{\tau}_t)$ and $\zeta(\tilde{\tau}_t)$, define the function $z(\tilde{\tau}_t)$ as follows:

$$z(\tilde{\tau}_t) = \begin{cases} \arg\max_{z_t \geq \tilde{\tau}_t} \{(1-\delta)w(z_t, \tilde{\tau}_t) + \delta w(z_t, z_t)\} & \text{if } \hat{\tau} > \zeta(\tilde{\tau}_t) \\ \arg\max_{z_t} \{(1-\delta)w(z_t, \tilde{\tau}_t) + \delta w(\hat{\tau}, \hat{\tau})\} & \text{if } \zeta(\tilde{\tau}_t) \geq \hat{\tau} \end{cases}. \tag{3.4}$$

The *optimal deviation function* $z(\tilde{\tau}_t)$ indicates how the optimal deviation, z_t, varies with the agreed tariff $\tilde{\tau}_t$. The next result provides a formal characterization of $z(\widetilde{\tau}_t)$:

Lemma 2. *Assume A1-A3. There is a unique solution to (3.3), for which* $\overline{\tau} < \widehat{\tau}$. *The solution to (3.4) satisfies: (i) for all* $\tilde{\tau}_t < \overline{\tau}$, $z(\tilde{\tau}_t) \geq \zeta(\tilde{\tau}_t) > \overline{\tau}$; *(ii) for all* $\tilde{\tau}_t \geq \overline{\tau}$, $z(\tilde{\tau}_t) = \tilde{\tau}_t$; *(iii)* $\lim_{\delta \to 0} \overline{\tau} = \hat{\tau}$ *and* $\lim_{\delta \to 1} \overline{\tau} = 0$.

Fig. 2 illustrates Lemma 2, showing how $\mu(\widetilde{\tau}_t)$ and $\zeta(\tilde{\tau}_t)$ are used to define $z(\tilde{\tau}_t)$, and how the inequality constraint $z_t \geq \tilde{\tau}_t$ applies. The solid line in Fig. 2 illustrates $z(\tilde{\tau}_t)$.

If, for some agreed tariff $\widetilde{\tau}_t < \overline{\tau}$, the optimal deviation given by $\zeta(\tilde{\tau}_t)$ is such that $\zeta(\tilde{\tau}_t) \geq \hat{\tau}$, then $z(\tilde{\tau}_t) = \mu(\tilde{\tau}_t)$. As explained above, for an initial deviation $z_t > \hat{\tau}$, WEC allows retaliation equal to z_t. But neither country can credibly commit to set tariffs greater than $\hat{\tau}$. So the continuation game has both countries setting $\hat{\tau}$ in all periods following the deviation. Thus, the optimal deviation is given by $\mu(\tilde{\tau}_t)$.

If, on the other hand, for some agreed tariff $\widetilde{\tau}_t < \overline{\tau}$ the optimal deviation given by $\zeta(\tilde{\tau}_t)$ is such that $\zeta(\tilde{\tau}_t) < \hat{\tau}$, then $z(\tilde{\tau}_t) = \zeta(\tilde{\tau}_t)$. For $z_t < \hat{\tau}$, the punisher *can* credibly commit to set its tariff equal to the initial deviation. In this case, a deviator takes into account that a higher deviation tariff brings about a higher punishment tariff, and therefore deviates by less than when the punishment is fixed at $\hat{\tau}$.[28] As a result, there

[28]For $z_t < \hat{\tau}$, a deviator deviates in accordance with $\zeta(\widetilde{\tau}_t)$ and not $\mu(\widetilde{\tau}_t)$, rationally anticipating that it will be punished in all future periods by z_t and not $\hat{\tau}$. Proposition 2 shows that, given sufficiently low discounting, following a deviation to $z_t < \hat{\tau}$, it is indeed rational to punish by setting z_t and not $\hat{\tau}$.

is a discontinuity in $z(\tilde{\tau}_t)$ at $\hat{\tau}$. Of course, it may be that $\hat{\tau} > \zeta(\tilde{\tau}_t)$ for all $\tilde{\tau}_t \in \mathbb{R}_+$, in which case $z(\tilde{\tau}_t) = \zeta(\tilde{\tau}_t)$ for all $\tilde{\tau}_t \in \mathbb{R}_+$. This becomes increasingly likely as $\delta \to 1$.

Note that $\zeta(\widetilde{\tau}_t) < \widetilde{\tau}_t$ for all $\widetilde{\tau}_t > \overline{\tau}$. But recall that WEC does not require downwards deviations to be matched; the constraint $z_t \geq \widetilde{\tau}_t$ in (3.4) is binding and so $z_t = \widetilde{\tau}_t$ for $\widetilde{\tau}_t > \overline{\tau}$.

Lemma 2 emphasizes that for all $\tilde{\tau}_t < \overline{\tau}$ the optimal deviation is to set $z(\tilde{\tau}_t) > \overline{\tau}$. This result will be useful in the characterization of the efficient equilibrium path (see Lemma 3 in particular). More broadly, for any $\widetilde{\tau}_t$ in a candidate equilibrium sequence $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$, we know from $z(\tilde{\tau}_t)$ the optimal deviation for that period under WEC.

Part (iii) of Lemma 2 indicates how $\overline{\tau}$ changes with δ. The result focuses on the extreme values of δ: $\delta \to 0$ and $\delta \to 1$. As mentioned above, $\zeta(\tilde{\tau}_t) \to \mu(\tilde{\tau}_t)$ in the limit as $\delta \to 0$, and so $\overline{\tau} \to \hat{\tau}$. At the other extreme, as $\delta \to 1$, the self enforcing tariff converges to free trade. Under the function $\zeta(\tilde{\tau}_t)$, a deviator puts an increasing weight on the payoff under future punishment in the continuation game and under WEC this is determined by the size of the initial deviation z_t. So as $\delta \to 1$, the optimal deviation converges to $z_t = 0$. This last result will be useful in showing that as $\delta \to 1$ the equilibrium tariff path converges towards free trade.

Before moving on to look at equilibrium paths, it remains to show that (3.1) generates a punishment path that is subgame perfect, and to establish conditions under which a withdrawal of equivalent concessions is itself a subgame perfect punishment strategy.

Proposition 1. *Assume A1-A3. The payoff given by (3.1) to deviation $z(\widetilde{\tau}_t)$ such that $\hat{\tau} > z(\widetilde{\tau}_t) = \zeta(\tilde{\tau}_t) > \tilde{\tau}_t$ generates a punishment path that is subgame perfect.*

Proposition 1 confirms that once a country has deviated to $z(\widetilde{\tau}_t) < \hat{\tau}$ and the other country has retaliated by also adopting $z(\widetilde{\tau}_t)$, neither country has an incentive to deviate again under WEC. At first sight it might appear that an initial deviation to $z(\widetilde{\tau}_t)$ could create the incentive for a further deviation to $z(z(\widetilde{\tau}_t)) > z(\widetilde{\tau}_t)$, once the punisher had also adopted $z(\widetilde{\tau}_t)$. In fact, this would never happen in the model of this present paper, and there is no incentive to deviate subsequently from $z(\widetilde{\tau}_t)$.

To see why there is no incentive to deviate subsequently from $z(\widetilde{\tau}_t)$, look at Fig. 3.

First note that in general an initial deviation may occur at $z(\widetilde{\tau}_t) > \overline{\tau}$, or $z(\widetilde{\tau}_t) = \overline{\tau}$. Consider the agreed tariff level $\widetilde{\tau}_t = \tau'$ shown in Fig. 3, from which there is an incentive to deviate initially to $z(\tau') > \overline{\tau}$. Now consider the incentive to deviate subsequently from $z(\tau')$. Fig. 3 shows that the optimal deviation from $z(\tau')$ is $z(\tau')$ itself.

Proposition 1 does not consider the case where $z(\widetilde{\tau}_t) = \mu(\tilde{\tau}_t) \geq \hat{\tau}$ because this case is well understood. As mentioned above, in this case the game corresponds precisely to a game of trigger strategies; while in the first period the deviator sets $z(\widetilde{\tau}_t)$ in response to $\widetilde{\tau}_t$ set by the other country, in the continuation game both countries adopt $\hat{\tau} < z(\widetilde{\tau}_t)$ and this is a subgame perfect equilibrium. Thus it is immediately obvious that the payoff given by the final line of (3.1) is also subgame perfect.

The next result establishes conditions under which the other country actually has an incentive to adopt a punishment consistent with WEC, given an initial deviation to $z(\tau) < \hat{\tau}$. The alternative would be to punish more severely than is allowed by WEC and trigger a suspension of GATT obligations under Article XXIII.

Proposition 2. *Assume A1, A2. Assume an initial deviation $z(\widetilde{\tau}_t)$ such that $\hat{\tau} > z(\widetilde{\tau}_t) = \zeta(\tilde{\tau}_t) > \widetilde{\tau}_t$. Then there is a $\delta' \in (0,1)$ such that, for all $\delta \in (\delta', 1)$, WEC is a subgame perfect punishment strategy.*

This result is based on Friedman's (1971) "Nash-threats" folk theorem. A country knows that if it punishes more aggressively than is allowed by the WEC rule, by setting $z' > z(\widetilde{\tau}_t)$ in response to an initial deviation $z(\widetilde{\tau}_t)$, then this will provoke a suspension of GATT obligations. As with the "Nash-threats" folk theorem, this triggers an indefinite imposition of tariffs at $\hat{\tau}$. On the other hand, if a punisher adheres to WEC, by setting $z(\widetilde{\tau}_t)$, then under the punishment phase tariffs are also set at $z(\widetilde{\tau}_t)$. And by A2, symmetric tariffs at $z(\widetilde{\tau}_t)$ yield a higher payoff than symmetric tariffs at $\hat{\tau}$. At the limit $\delta = 1$, countries are hurt more by the lower payoff under suspension of GATT obligations $\hat{\tau}$. So there must exist a range $\delta \in (\delta', 1)$ for which countries prefer to punish by WEC. For δ outside the interval $(\delta', 1)$, a punisher prefers to punish using standard trigger strategies, given that an initial deviation has taken place.

It is worth mentioning that for the tariff game specified in this present section, in which only the actions of punishers are limited by WEC, there is a unique efficient equilibrium tariff liberalization path with $\widetilde{\tau}_0 = \hat{\tau}$ and $\widetilde{\tau}_t = \overline{\tau}$ for all $t \geq 1$. That is, some trade liberalization can be achieved (moving from $\hat{\tau}$ to $\overline{\tau}$) but free trade cannot be reached and the efficient equilibrium path does not exhibit gradualism. (This equilibrium tariff path is derived formally in Lockwood and Zissimos 2002, while the present paper focuses on gradualism.) This contrasts markedly with the equilibrium path that we will derive in the next section, in which $\widetilde{\tau}_0 = \hat{\tau}$ but $\widetilde{\tau}_1 < \overline{\tau}$, and tariffs are strictly decreasing thereafter, thus getting closer to free trade (although free trade is not reached in finite time).[29]

4. Gradualism

This section characterizes the equilibrium tariff reduction path. We will use the analysis of the previous section, which characterized limited punishments under WEC. In order to isolate and understand the effects of WEC, our analysis in the previous section imposed no restrictions on deviations. Thus we derived an optimal deviation, z_t. However, following the previous literature in this area I now argue that the actions not just of punishers but of initial deviators are likely to be limited as well under WEC. Specifically, the institutional feature of the GATT that we will now introduce is that a deviator cannot expect to be protected by WEC if its deviation is deemed to be abusive. In this section, we will define formally what constitutes an abusive deviation, and use this definition to introduce limits on the actions of deviators (while limits on punishers in the form of WEC are retained). Equilibrium is then characterized in the two main results of this paper. (In Appendix A.2 there is an extended discussion relating the model of this present paper to the literature.)

The question we need to address now is how much can a government deviate before the offending action is deemed to be abusive? As pointed out in the Introduction, there is no basis written into GATT Articles on which to discriminate between a deviation that

[29] Lockwood and Zissimos (2002) simply assume that $\zeta(\widetilde{\tau}_t) < \hat{\tau}$ and do not identify the possibility that, for any $\widetilde{\tau}_t$ such that $\zeta(\widetilde{\tau}_t) > \hat{\tau}$,

$$(1-\delta) w(\zeta(\widetilde{\tau}_t), \widetilde{\tau}_t) + \delta w(\zeta(\widetilde{\tau}_t), \zeta(\widetilde{\tau}_t)) < (1-\delta) w(\mu(\widetilde{\tau}_t), \widetilde{\tau}_t) + \delta w(\hat{\tau}, \hat{\tau}).$$

But it follows from Proposition 2 that, for $\delta \in (\delta', 1)$ identified in Proposition 2, it must be the case that $\zeta(\widetilde{\tau}_t) < \hat{\tau}$ for all $\widetilde{\tau}_t$. This holds because, by A2, $w(\tau, \tau)$ is decreasing in τ. So the region for which $\zeta(\widetilde{\tau}_t) > \hat{\tau}$ is not relevant if $\delta \in (\delta', 1)$. Therefore, for $\delta \in (\delta', 1)$ the analysis presented by Lockwood and Zissimos is consistent with the analytical framework and results of this present paper. I am grateful to a referee for bringing this point to light.

brings about a withdrawal of equivalent concessions and one that provokes a suspension of GATT obligations. Since no formalization of such circumstances exists in GATT rules, we must introduce an interpretive assumption about what affected parties will tolerate.

To do this, say that along the candidate equilibrium path, at period $t-1$, a given tariff level $\widetilde{\tau}_{t-1}$ is achieved. The aim is then to achieve some lower tariff level $\widetilde{\tau}_t$. At period $t-1$, the target tariff level $\widetilde{\tau}_t$ for period t is then said to be the *binding* for period t or *scheduled binding* (Article II).[30] By contrast, bindings already achieved for $t-s$, $s \geq 1$ are said to be *past bindings.*

Country i *breaks a scheduled binding* by setting $\tau_t = \tau'$ such that $\tau' > \widetilde{\tau}_t$; in period t it fails to reduce its tariff to $\widetilde{\tau}_t$. The punishment imposed on country i by country j then depends on the extent to which the binding is broken. There are two possibilities.

(i) In period t, country i *breaks a scheduled binding* but does not break a past binding; it fails to reduce its tariff to the scheduled level but it does not raise the tariff above the level achieved in $t-1$; $\widetilde{\tau}_{t-1} \geq \tau_t = \tau' > \widetilde{\tau}_t$. Then in period $t+1$ country j withdraws equivalent concessions.

(ii) In period t, country i *breaks a past binding* by raising its tariff above the level set in $t-1$; $\tau_t = \tau' > \widetilde{\tau}_{t-1}$. Then in all future periods there is an indefinite suspension of GATT obligations; country j sets $\hat{\tau}$ from $t+1$ onwards.[31]

We will say that when the actions of initial deviators are restricted in this way and the actions of punishers are restricted by WEC as described in Section 3 then we have a

[30] In practice, a tariff binding has come to be understood to represent one of three types of commitment: (1) to lower a tariff (duty) to a stated level; (2) not to raise a tariff above its current level; (3) not to raise a tariff above a specified higher level (Dam 1970, p. 31). In this present paper we focus on the most popular usage of the term binding given by (1).

[31] Given that such a distinction is called for by Bagwell and Staiger (2002), drawing the line at a past binding seems as reasonable as any. And it is entirely consistent with the situation formalized by Bagwell and Staiger (1990, 2002) where a large but verifiable shock causes a past binding to be broken but is nevertheless redressed by WEC.

It may nevertheless be argued that in practice parties have tolerated a breach of past bindings, and responded by WEC. The present model could be extended to allow a past binding to be broken by a certain margin $\lambda\tau_{t-1}$, where $\lambda > 1$, provided that $\lambda\tau_{t-1} << \hat{\tau}$. However, the analysis on which this extension is based is more complex, without changing the gradualism result qualitatively. Given, by Lemma 2, that all $z(\tau) > \overline{\tau}$ for $\tau < \overline{\tau}$, for the extension we would need an analogous $\vec{\tau}$ (say) for which $z(\vec{\tau}) = \lambda\vec{\tau}$. By construction it must be the case that $\vec{\tau} < \overline{\tau}$. Then $z(\tau) > \vec{\tau}$ for $\tau < \vec{\tau}$. Because deviation from $\vec{\tau}$ must be more attractive than deviation from $\overline{\tau}$, there must be an upper bound on the value of λ and the range of feasible $\delta < 1$ must be reduced.

tariff game with bindings.

To summarize, the interpretive assumption introduced here essentially implies that, even if it is not backed by a stochastic shock, an 'opportunistically motivated' deviation is taken in good faith and is simply matched by a WEC as long as it does not break a past binding.

We can now show that if countries are sufficiently patient then they will not break a past binding under deviation, in the knowledge that doing so would bring about a suspension of GATT obligations.

Proposition 3. *Assume A1-A3 and that governments play a tariff game with bindings. Then there is a* $\underline{\delta} \in [\delta', 1)$ *such that for all* $\delta \in (\underline{\delta}, 1)$ *if a country deviates from a candidate equilibrium path* $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$ *it does not break a past binding.*

As with Proposition 2, this result is based on Friedman's (1971) "Nash-threats" folk theorem. A country knows that if it deviates from a candidate equilibrium path in such a way that it breaks a past binding, by setting $\tau_t = \tau' > \widetilde{\tau}_{t-1}$, then this will provoke a suspension of GATT obligations; an indefinite imposition of tariffs at $\hat{\tau}$. On the other hand if a country deviates in such a way that it does not break a past binding, by setting $\tau' = \widetilde{\tau}_{t-1}$, then under the punishment phase tariffs are $\widetilde{\tau}_{t-1}$. (By Proposition 2, this holds only for $\delta \in (\delta', 1)$. Outside of this range, punishment is by trigger strategies.) And by A2, symmetric tariffs at $\widetilde{\tau}_{t-1}$ yield a higher payoff in the continuation game than symmetric tariffs at $\hat{\tau}$. At the limit $\delta = 1$, countries are hurt more by the lower payoff under a suspension of GATT obligations $\hat{\tau}$. So there must exist a range of $\delta < 1$ for which countries do not break past bindings under deviation.[32]

The proposition does not hold for δ outside the interval $(\delta', 1)$. In that case, by Proposition 2, punishers fail to adhere to WEC and instead trigger a suspension of GATT

[32]The interpretive assumption, resulting in the selection of $\tau' = \widetilde{\tau}_{t-1}$, can be thought of as a 'refinement' for the off-equilibrium tariff paths of the game, the purpose of which is to introduce a degree of irreversibility into the initial deviation. This is to be distinguished from the usual equilibrium refinements which eliminate certain equilibrium tariff paths. Building on the insights of Lockwood and Thomas (2002), the assumption that the off-equilibrium tariff paths imply partial irreversibility (as opposed to a return all the way to trigger strategies) is key to motivating gradualism as a result. One would not expect to be able to generate gradualism in a non-trivial way by the usual refinements of the equilibrium path.

obligations, punishing by using standard trigger strategies for any deviation from the equilibrium path. Knowing that it will be punished by trigger strategies, a deviator has no incentive to deviate in a limited way and would deviate according to $\mu(\widetilde{\tau}_t)$. Moreover, Proposition 3 does not necessarily hold over the entire range for which Proposition 2 holds. We cannot rule out the possibility that there exists an interval $\delta \in (\delta', \underline{\delta}]$ for which a deviator prefers to deviate by setting $\tau' > \widetilde{\tau}_{t-1}$, triggering a suspension of GATT obligations.

How much does the bound on δ limit the applicability of Proposition 3 and the others that rely on it? It is generally recognized that as long as agents are able to adjust the interval between periods, they can ensure that δ is in the required range for the result to hold when it is in their interest. (See Scherer 1980 and Fudenberg and Tirole 1991 for further discussion.) So a consistent explanation of the fact that past bindings were rarely broken (until 1994) is that dispute resolution procedures operated sufficiently quickly.

4.1. Efficient equilibrium tariff reduction paths

We will now formally define the conditions that must hold if a symmetric tariff path is to be a subgame-perfect one in our game. In every period, the continuation payoff from the path must be at least as great as the maximal payoff from deviation, given that a punishment consistent with WEC will ensue. From Lemma 1, the maximal payoff from an optimal deviation at t is $(1-\delta)w(z(\widetilde{\tau}_t), \widetilde{\tau}_t) + \delta w(z(\widetilde{\tau}_t), (z(\widetilde{\tau}_t))$. But under a tariff game with bindings, and with $\delta \in (\underline{\delta}, 1)$, in the event that a country deviates the "optimal deviation" given in (3.4) is not chosen unless $z(\widetilde{\tau}_t) \leq \widetilde{\tau}_{t-1}$. Otherwise, $\widetilde{\tau}_{t-1}$ will be chosen under deviation. We will establish in this subsection the conditions under which a deviation must be equal to $\widetilde{\tau}_{t-1}$, facilitating a simple characterization of the equilibrium path.

Defining $\chi(\widetilde{\tau}_t, \widetilde{\tau}_{t-1}) = \min\{z(\widetilde{\tau}_t), \widetilde{\tau}_{t-1}\}$, the equilibrium condition is

$$\begin{aligned} (1-\delta)(w(\widetilde{\tau}_t, \widetilde{\tau}_t) + \delta w(\widetilde{\tau}_{t+1}, \widetilde{\tau}_{t+1}) + \ldots \quad) &\geq \\ (1-\delta)w(\chi(\widetilde{\tau}_t, \widetilde{\tau}_{t-1}), \widetilde{\tau}_t) + \delta w(\chi(\widetilde{\tau}_t, \widetilde{\tau}_{t-1}), \chi(\widetilde{\tau}_t, \widetilde{\tau}_{t-1})),\ t &= 1, \ldots \end{aligned} \tag{4.1}$$

The left hand side gives the current discounted payoff to symmetrical tariff reductions while the right hand side gives the discounted payoff to a deviation. Of course, a whole set of paths will satisfy this sequence of inequalities: let this set of equilibrium paths be denoted E. An *efficient tariff reduction path* in the set E is simply a sequence $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$ of tariffs in E for which there is no other sequence $\{\widetilde{\tau}'_t\}_{t=1}^{\infty}$ also in E which gives a higher payoff to any country, as calculated by (2.1). Following the arguments of Lockwood and Thomas (2002), it can be shown that if $\{\widetilde{\tau}_t\}_{t=1}^{\infty}$ is efficient, (4.1) holds with equality at every date i.e.:

$$\begin{aligned}(1-\delta)(w(\widetilde{\tau}_t,\widetilde{\tau}_t)+\delta w(\widetilde{\tau}_{t+1},\widetilde{\tau}_{t+1})+...\quad) &= \\ (1-\delta)w(\chi(\widetilde{\tau}_t,\widetilde{\tau}_{t-1}),\widetilde{\tau}_t)+\delta w(\chi(\widetilde{\tau}_t,\widetilde{\tau}_{t-1}),\chi(\widetilde{\tau}_t,\widetilde{\tau}_{t-1})),\ t &= 1,...\end{aligned} \tag{4.2}$$

The intuition is that if (4.1) held with strict inequality, it would be possible to reduce the tariff path by a small amount without violating (4.1).

Of the class of equilibrium paths E, it is obviously the set of efficient paths that is of most interest. It is generally accepted that the GATT provides a mechanism through which countries are able to coordinate their selection of an efficient equilibrium path (Bagwell and Staiger 1990). We now turn to a characterization of the set of efficient equilibrium paths.

Proposition 3 says that there exists a range of δ for which a country's deviation will not exceed $\widetilde{\tau}_{t-1}$. We now establish the conditions under which a country actually has an incentive to set a tariff under deviation that is as high as $\widetilde{\tau}_{t-1}$ (or higher, if this did not break a past binding and provoke a suspension of GATT obligations).

Lemma 3. *Assume A1-A3, let $\delta \in (\underline{\delta}, 1)$, and let $\overline{\tau}$ be the (unique) value for which $z(\tau) = \zeta(\tau) = \tau$ given δ. If $\widetilde{\tau}_t, \widetilde{\tau}_{t-1} \leq \overline{\tau}$, then $\chi(\widetilde{\tau}_t, \widetilde{\tau}_{t-1}) = \widetilde{\tau}_{t-1}$.*

The thinking behind Lemma 3 is illustrated in Fig. 4.[33] From the figure we see that for any $\widetilde{\tau}_t, \widetilde{\tau}_{t-1} \leq \overline{\tau}$, the previous period tariff is less than the optimal deviation; $\widetilde{\tau}_{t-1} < z(\widetilde{\tau}_t)$. If a country could deviate freely then it would deviate to $z(\widetilde{\tau}_t)$, but in a tariff game with bindings it can only deviate to $\widetilde{\tau}_{t-1}$.

From Lemma 3, for all t in which $\widetilde{\tau}_t, \widetilde{\tau}_{t-1} \leq \overline{\tau}$, the previous period tariff $\widetilde{\tau}_{t-1}$ can be substituted for $\chi(\widetilde{\tau}_t, \widetilde{\tau}_{t-1})$ in the equilibrium condition (4.2), facilitating a simple characterization of an efficient equilibrium tariff reduction path.

[33] In Figure 4, $\delta \in (\underline{\delta}, 1)$ so $z(\widetilde{\tau}_t) < \hat{\tau}$ for all $\widetilde{\tau}_t$; see Footnote 29.

It takes two steps to prove that an efficient equilibrium tariff reduction path is gradually decreasing under the tariff game with bindings. Initially we will simply assume, in Lemma 4, that $\widetilde{\tau}_t, \widetilde{\tau}_{t-1} \leq \overline{\tau}$ for all $t \geq 1$ and show that an efficient equilibrium tariff sequence must be strictly decreasing. Then in the first main result of the paper, Proposition 4, we will remove the assumption that $\widetilde{\tau}_t, \widetilde{\tau}_{t-1} \leq \overline{\tau}$ and prove that for an efficient equilibrium tariff sequence it must be the case that $\widetilde{\tau}_1 < \overline{\tau}$. Gradual tariff reductions then follow by Lemma 4. Let us first see the results then consider the intuition.

From (4.2) and Lemma 3, an efficient equilibrium path with $\widetilde{\tau}_t \leq \overline{\tau}$, $t \geq 1$ must satisfy[34]

$$w(\widetilde{\tau}_t, \widetilde{\tau}_{t+1}) = \frac{1}{\delta}\left[w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_t) - w(\widetilde{\tau}_t, \widetilde{\tau}_t)\right] + \frac{w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_{t-1})}{1-\delta} - \frac{\delta w(\widetilde{\tau}_t, \widetilde{\tau}_t)}{1-\delta}, \; t > 1. \tag{4.3}$$

Let $\{\tau_t(\widetilde{\tau}_0, \widetilde{\tau}_1)\}_{t=2}^{\infty}$ be the sequence that solves (4.3) with initial conditions $\widetilde{\tau}_0, \widetilde{\tau}_1$. We can now establish gradualism by showing that as long as there is a tariff reduction in the first period then tariffs must strictly fall in all subsequent periods along any efficient equilibrium path.

Lemma 4. *Assume A1 and $\delta \in (\underline{\delta}, 1)$. Any sequence $\{\tau_t(\widetilde{\tau}_0, \widetilde{\tau}_1)\}_{t=2}^{\infty}$ that satisfies (4.3), with initial conditions $\widetilde{\tau}_0, \widetilde{\tau}_1$ that satisfy $0 < \widetilde{\tau}_1 < \widetilde{\tau}_0$ and $0 < \widetilde{\tau}_1 < \overline{\tau}$, is strictly decreasing i.e. $0 < \tau_{t+1}(\widetilde{\tau}_0, \widetilde{\tau}_1) < \tau_t(\widetilde{\tau}_0, \widetilde{\tau}_1)$ all $t \geq 2$.*

Now consider the construction of an efficient path, given this result. First, $\widetilde{\tau}_0$ is fixed at $\hat{\tau}$. Second, from $t = 2$ onwards, i.e. conditional on $\widetilde{\tau}_0, \widetilde{\tau}_1$, an efficient path is simply $\{\tau_t(\widetilde{\tau}_0, \widetilde{\tau}_1)\}_{t=2}^{\infty}$ as long as (i) $\widetilde{\tau}_1 < \widetilde{\tau}_0$ (required by Lemma 4), and (ii) $\widetilde{\tau}_1 \leq \overline{\tau}$ (required by Lemma 3; otherwise, the efficient path in question does not satisfy (4.3)). So, it remains to show that on the efficient equilibrium path, $\widetilde{\tau}_1$ will be chosen to satisfy $\widetilde{\tau}_1 \leq \overline{\tau} < \hat{\tau}$. If a path is to be efficient, the incentive constraint (4.1) must hold with equality at $t = 1$, i.e.

$$\begin{aligned} & (1-\delta)(w(\widetilde{\tau}_1, \widetilde{\tau}_1) + \delta w(\widetilde{\tau}_2(\hat{\tau}, \widetilde{\tau}_1), \widetilde{\tau}_2(\hat{\tau}, \widetilde{\tau}_1)) + ...) \\ = \; & (1-\delta) w(\chi(\widetilde{\tau}_1, \hat{\tau}), \widetilde{\tau}_1) + \delta w(\chi(\widetilde{\tau}_1, \hat{\tau}), \chi(\widetilde{\tau}_1, \hat{\tau})) \end{aligned} \tag{4.4}$$

We can now establish the first main result, that in a tariff game with bindings trade liberalization must be gradual.

[34] See the proof of Lemma 4 for a full derivation of (4.3).

Proposition 4. *Assume A1-A3, $\delta \in (\underline{\delta}, 1)$, and that governments play a tariff game with bindings. There exists a smallest value of $\tilde{\tau}_1$, $0 < \tilde{\tau}_1 < \overline{\tau}$, that satisfies (4.4). Consequently, there exists an efficient equilibrium path $(\tilde{\tau}_1, \tilde{\tau}_2, \tilde{\tau}_3, ...)$ with $\tilde{\tau}_t = \tau_t(\hat{\tau}, \tilde{\tau}_1)$, $t > 1$. This path exhibits a gradually decreasing tariff i.e. $\tilde{\tau}_{t+1} < \tilde{\tau}_t$, $t \geq 1$.*

From Proposition 4 we learn that under a tariff game with bindings it *is* possible to achieve a gradually declining equilibrium path for which $\tilde{\tau}_t < \overline{\tau}$, all $t \geq 1$. Consider some period s in which tariffs have been reduced by a gradual process over periods $t = 1, ..., s$ to some tariff level $\tilde{\tau}_s < \overline{\tau}$. Now suppose that the agreement requires $\tilde{\tau}_{s+1} < \tilde{\tau}_s$ in period $s+1$. If the agreement proposes no further reductions in periods after that, then country j does better by deviating while country i proceeds to set $\tilde{\tau}_{s+1} < \tilde{\tau}_s$, even if country i imposes WEC in all periods after that. But by Proposition 3, country j does not break a past binding under deviation; its optimal deviation is to maintain $\tilde{\tau}_s$ in $s+1$.[35] Because punishment is limited, current tariff cuts can only be made self-enforcing by the promise of future tariff reductions. And by Lemma 4, given limited deviations, it is always possible for country i to promise additional reductions in future periods that can more than compensate for the gains to deviation in period $s+1$. This is gradualism in other words.

The basic intuition of this gradualism result is as follows. Gradualism arises in the presence of WEC because, by pursuing a gradual liberalization path, countries can limit the incentive to deviate at any point in time to a reversion up to the previous period's tariff binding. This is because they are only protected by the WEC rule up to that point. And this diminished ability to deviate at any point in time, combined with the continuing inducements of further liberalization at all points in the future, allows governments to ultimately reach lower tariff levels than they would be able to under once-for-all immediate liberalization from the static Nash equilibrium tariff rate.

The second main result is that in a tariff game with bindings free trade cannot be reached.

[35] This is conditional upon $\delta \in (\underline{\delta}, 1)$.

Proposition 5. *(Failure to Reach Free Trade). Assume A1-A3,* $\delta \in (\underline{\delta}, 1)$*, and that governments play a tariff game with bindings. Then the value* $\tilde{\tau}_t = 0$ *cannot be reached on an efficient equilibrium path* $\tau_t(\hat{\tau}, \tilde{\tau}_1)$*,* $t > 1$*. However, as* $\delta \to 1$*, any efficient path converges to free trade.*

This result follows naturally when it is realized that, not only is it possible to promise reductions in future periods that can more than compensate for gains to deviation today, it is actually necessary to promise further liberalization in future periods on the equilibrium path because the ability to sustain an agreement through punishment is limited. If there are no reductions scheduled for the future then there is no incentive to implement agreed reductions today. This applies as much to the scheduling of free trade as to any other proposed floor of the tariff liberalization path.

As countries are made extremely patient ($\delta \to 1$) it is possible to move almost all the way to free trade in the first period of trade liberalization and then continue to move gradually towards free trade in very small steps. Recall from Lemma 2(iii) that $\lim_{\delta \to 1} \overline{\tau} = 0$. And recall from Proposition 4 that on the equilibrium path $\widetilde{\tau}_1 < \overline{\tau}$, $\widetilde{\tau}_{t+1} < \widetilde{\tau}_t$. Even in this case free trade will not be reached in finite time.[36] As we move away from such extreme levels of patience we should not expect such convergence towards free trade to be preserved. Intuitively, the lower the level of δ the higher the payoff to a deviation that brings about a suspension of GATT obligations. This puts a lower bound on the tariff liberalization that can be achieved on an efficient equilibrium path.[37]

5. Conclusions

This paper helps to explain two stylized facts about trade liberalization, namely gradualism and failure to reach free trade. It does so by studying the interplay between countries' unilateral incentive to set tariffs and aspects of the institutional structure set up in the framework of the GATT to achieve trade liberalization. In particular, aspects

[36] By the continuity of $w(\cdot, \cdot)$, it is always possible to choose $\widetilde{\tau}_1$ such that $0 < \widetilde{\tau}_1 < \overline{\tau}$ even though $\lim_{\delta \to 1} \overline{\tau} = 0$.

[37] I am not able to formalize this intuition without imposing additional structure on the model. However, simulations based on quasi-linear preferences bear out the intuition; the tariff liberalization paths appear to converge towards a finite positive limit, and this limit is decreasing in δ; see Lockwood and Zissimos (2002).

of the GATT's Articles are shown to impose partial irreversibility on countries' ability to set tariffs. First, a withdrawal of equivalent concessions limits the ability of contracting parties to raise tariffs in order to punish a country for deviating from an agreement under Article XXIII. Second, while relatively small deviations may be punished by a withdrawal of equivalent concessions, a larger deviation will provoke a suspension of GATT obligations. (As explained above, this second feature is conditional upon an interpretive assumption about parties' tolerance of deviations and how they are assumed to use available GATT rules as a result.)

The starting point of the analysis is to take the GATT rules that we study as given and assume that two countries have each already signed the GATT. We then play out a dynamic tariff reduction game and characterize the equilibrium path. Off-equilibrium-path play is fully characterized by drawing on aspects of Articles XIX and XXIII. In so doing, we are able to show that (sufficiently patient) countries can do no better than to keep to a symmetrical gradual tariff reduction path.

We have shown that once countries have signed up to the GATT, they can do no better than to liberalize according to the rules that it sets out. Yet it is clear that the GATT would achieve greater efficiency if it sanctioned more severe punishments of deviators. This raises an important question for future research, namely, what features are omitted from the model in this present paper that might make WEC and the possibility to deviate under Articles XIX and XXIII achieve efficiency? One direction in which to pursue an answer would be to introduce stochastic productivity shocks to our model. Because it is deterministic, our model rules out the possibility that Article XIX will actually be used, as it was originally intended, to relax the participation constraint in a trade agreement in the face of shocks that would otherwise precipitate the agreement's break-up and a return to greater protectionism. While the analysis shows that WEC in conjunction with Article XIX prevents efficiency from being achieved, in a stochastic world WEC might be efficiency enhancing by enabling an agreement to survive. The efficiency enhancing potential of Articles XIX and XXIII is discussed by Bagwell and Staiger (1990, and 2002 Chapter 6). The interplay of these competing roles for deviation and retaliation presents an interesting agenda for future research.

Inevitably, the theoretical framework developed here simplifies the situation in a number of other key respects. Countries are assumed to be symmetrical, each country exports only a single good, with both countries equally open at a given time. In practice countries export a number of goods, with levels of openness varying across sectors. Variation in country size and purchasing power across different markets is likely to make the actual dynamics of liberalization considerably more subtle and complex. Gradualism in a context where there are asymmetries across countries has been studied by Bond and Park (2002), but not within the context of the GATT penalty structure that we examine here.

A promising direction for future research would allow trade block formation to be considered. The theory of repeated games has been used to study trade block formation, where a preferential trade agreement is supported by the credible threat of punishment. In their paper, which uses a repeated game framework, Bond, Syropoulos and Winters (2001) point out that trade liberalization within the European Union has been very slow.[38] It may be that the framework of this present paper could be adapted to provide a way of understanding gradualism between members.

There may be many other competing pressures other than the standard terms-of-trade motive working against further liberalization, and these are also suppressed in the present model. One area that has attracted significant attention recently is the incentive for politicians to give in to protectionist inducements from interest groups (Grossman and Helpman 1995). These protectionist forces may have been outweighed at an early stage in the post-war trade liberalization process when liberalization gains were large relative to the rents from protectionism, but not later once the potential trade gains began to be exhausted. Future research could usefully study the interaction of these counteracting forces.

The main point of the present paper is that under GATT rules trade liberalization must be gradual. A natural question follows as to 'how gradual' trade liberalization becomes as a result. One possible way forward is to use simulations. As mentioned above, Lockwood and Zissimos (2002) make some progress in this direction by undertaking simulations based on quasi-linear preferences. Yet these were unsatisfactory in that they appeared to show 'too much liberalization too soon' given what we have actually observed.

[38] Bond, Riezman and Syropoulos (2004) draw attention to the gradual nature of liberalization within FTAs.

Further work is needed to establish how the forces for gradualism vary with different functional forms and whether gradualism is exacerbated by other forces for protectionism such as lobbying.

A. Appendix

A.1. Proof of propositions

Proof of Lemma 1. (a) First, suppose that a country deviates to $z_t < \widetilde{\tau}_t$. Then, from (3.1), as there is no retaliation, future payoffs are unaffected by the choice of deviation. Moreover, as $w(z_t, \tilde{\tau}_t)$ is increasing in z_t by A1, the payoff to deviation of the form $z_t < \widetilde{\tau}_t$ is increasing in z_t. Therefore, the supremum of the payoff to this kind of deviation is

$$\lim_{z_t \to \widetilde{\tau}_t} [(1-\delta)w(z_t, \tilde{\tau}_t) + (1-\delta) \sum_{s=t+1}^{\infty} \delta^{s-t} w(\tilde{\tau}_s, \tilde{\tau}_s)] = (1-\delta) \sum_{s=t}^{\infty} \delta^{s-t} w(\tilde{\tau}_s, \tilde{\tau}_s)$$

(b) If a country deviates to $z_t > \widetilde{\tau}_t$, it receives

$$g(z_t, \tilde{\tau}_t) = (1-\delta)w(z_t, \tilde{\tau}_t) + \delta w(\min\{z_t, \hat{\tau}\}, \min\{z_t, \hat{\tau}\}) \tag{A.1}$$

So, it suffices to show that (A.1) has a global maximum z_t^* on $(\tilde{\tau}_t, \infty)$. If this is *not* the case, then there exists an increasing sequence $\{z^n\}$ with $\lim_{n\to\infty} z^n \to \infty$, for which $g(z^n, \tilde{\tau}_t)$ is monotonically increasing. But for z^n high enough, $g(z^n, \tilde{\tau}_t)$ is decreasing in z^n; by A1, $\lim_{\tau\to\infty} w_1(\tau, \tau') < 0$ and so $(1-\delta)w(z_t, \tilde{\tau}_t)$ is decreasing in z^n as $\lim_{n\to\infty} z^n \to \infty$, and $\delta w(\min\{z_t, \hat{\tau}\}, \min\{z_t, \hat{\tau}\}))$ is fixed in z^n because $\min\{z_t, \hat{\tau}\} = \hat{\tau}$ as $\lim_{n\to\infty} z^n \to \infty$. Contradiction. □

Proof of Lemma 2. As $\zeta(\tilde{\tau}_t)$ is strictly decreasing and continuous in $\tilde{\tau}_t$, it must be the case that there exists a unique $\overline{\tau}$ for which $\zeta(\tilde{\tau}_t) > \tilde{\tau}_t$, given $\tilde{\tau}_t < \overline{\tau}$, and $\zeta(\tilde{\tau}_t) < \tilde{\tau}_t$ for $\tilde{\tau}_t > \overline{\tau}$.

We now prove that $\overline{\tau} < \hat{\tau}$. Suppose not; consider $\overline{\tau} = \hat{\tau}$ first. By the definition of (3.2) we must have $\zeta(\hat{\tau}) = \hat{\tau} = \arg\max_{z_t} \{(1-\delta)\, w(\hat{\tau}, \hat{\tau}) + \delta w(\hat{\tau}, \hat{\tau})\}$. The first order condition requires that

$$(1-\delta)\, w_1(\hat{\tau}, \hat{\tau}) + \delta(w_1(\hat{\tau}, \hat{\tau}) + w_2(\hat{\tau}, \hat{\tau})) = 0.$$

By a standard argument (using A1 and A3), $\hat{\tau}$ solves $w_1(\hat{\tau},\hat{\tau}) = 0$. But by A2, we have that $w_1(\hat{\tau},\hat{\tau}) + w_2(\hat{\tau},\hat{\tau}) < 0$. Therefore, the first order condition cannot be satisfied at $\overline{\tau} = \hat{\tau}$; a contradiction. Then $\overline{\tau} > \hat{\tau}$ can also be ruled out because $w_1(\overline{\tau},\overline{\tau}) < 0$ for $\overline{\tau} > \hat{\tau}$.

(i) For $\tilde{\tau}_t < \overline{\tau}$, by (3.4), if $\zeta(\tilde{\tau}_t) < \hat{\tau}$ then $z(\tilde{\tau}_t) = \zeta(\tilde{\tau}_t)$ and if $\zeta(\tilde{\tau}_t) \geq \hat{\tau}$ then $z(\widetilde{\tau}_t) = \mu(\widetilde{\tau}_t) \geq \hat{\tau}$. We have already shown that $\zeta(\tilde{\tau}_t) > \overline{\tau}$ for $\tilde{\tau}_t < \overline{\tau}$. It is then sufficient to show that $\mu(\widetilde{\tau}_t)$ lies everywhere above $\zeta(\tilde{\tau}_t)$. By definition, $\mu(\widetilde{\tau}_t)$ gives the value of z_t that solves

$$(1-\delta)\, w_1(z_t, \widetilde{\tau}_t) = 0.$$

We know by A1 that such a value of z_t exists. Also by definition, $\zeta(\widetilde{\tau}_t)$ gives the value of z_t that solves

$$(1-\delta)\, w_1(z_t, \widetilde{\tau}_t) + \delta\,(w_1(z_t, z_t) + w_2(z_t, z_t)) = 0.$$

By A1, $w_1(z_t, \widetilde{\tau}_t) > 0$ for $z_t < \mu(\widetilde{\tau}_t)$ and by A2 $w_1(z_t, z_t) + w_2(z_t, z_t) < 0$ for $z_t > 0$. It follows that $\zeta(\widetilde{\tau}_t) < \mu(\widetilde{\tau}_t)$ for all $\widetilde{\tau}_t$.

(ii) For $\tilde{\tau}_t > \overline{\tau}$, in the absence of the inequality constraint $z_t \geq \tilde{\tau}_t$ in (3.4) we would have $z(\widetilde{\tau}_t) = \zeta(\tilde{\tau}_t) < \tilde{\tau}_t$. But $z_t \geq \tilde{\tau}_t$ is binding and so $z(\widetilde{\tau}_t) = \tilde{\tau}_t$. Obviously, $z(\widetilde{\tau}_t) = \overline{\tau}$ for $\tilde{\tau}_t = \overline{\tau}$.

(iii) By the definition of (3.2) we must have $\zeta(\widetilde{\tau}_t) = \arg\max_{z_t} \{(1-\delta)\, w(z_t, \widetilde{\tau}_t) + \delta w(z_t, z_t)\}$. The first order condition requires that

$$(1-\delta)\, w_1(z_t, \widetilde{\tau}_t) + \delta(w_1(z_t, z_t) + w_2(z_t, z_t)) = 0.$$

In the limit as $\delta \to 0$, the first order condition tends towards the first order condition of the static tariff game; $w_1(z_t, \widetilde{\tau}_t) = 0$. Since $\overline{\tau}$ is the symmetric solution, we require the symmetric self-enforcing tariff of the static game, which is of course $\hat{\tau}$; it follows that $\lim_{\delta\to 0} \overline{\tau} = \hat{\tau}$.

In the limit as $\delta \to 1$, the first order condition tends towards the condition for efficient trade; $w_1(z_t, z_t) + w_2(z_t, z_t) = 0$. By A2, the solution is given by $z_t = 0$; it follows that $\lim_{\delta\to 1} \overline{\tau} = 0$. □

Proof of Proposition 1. Suppose not. Suppose instead that after country i initially deviates in period t to z_t and country j matches the deviation in $t+1$ by also setting

z_t, there is an incentive for some country to deviate from z_t in some period $s > t+1$. By Lemma 2, the initial deviation has two characterizations: (i) if $\widetilde{\tau}_t < \overline{\tau}$ then $z(\tilde{\tau}_t) \geq \zeta(\tilde{\tau}_t) > \overline{\tau}$; (ii) if $\widetilde{\tau}_t \geq \overline{\tau}$ then $z(\widetilde{\tau}_t) = \widetilde{\tau}_t$. Each will be taken in turn.

(i) If $\widetilde{\tau}_t < \overline{\tau}$ and so $z(\tilde{\tau}_t) = \zeta(\tilde{\tau}_t) > \overline{\tau}$ then, by Lemma 2 (ii), the best response to $z(\widetilde{\tau}_t)$ must also be $z(\widetilde{\tau}_t)$. To see why, observe that for $\zeta(\widetilde{\tau}_t) > \overline{\tau}$ the unconstrained best response given by (3.2) is $\zeta(\widetilde{\tau}_t) < \overline{\tau}$ but that the constraint $z_t \geq \widetilde{\tau}_t$ in (3.4) is binding. Therefore, $z(\widetilde{\tau}_t)$ is a best response to itself; contradiction. Of course, $\overline{\tau}$ is self enforcing, so the contradiction is immediate.

(ii) If $\widetilde{\tau}_t \geq \overline{\tau}$ then $z(\widetilde{\tau}_t) = \widetilde{\tau}_t \geq \overline{\tau}$ and so, again by Lemma 2 (ii), the best response to $z(\widetilde{\tau}_t)$ must also be $z(\widetilde{\tau}_t)$; contradiction. □

Proof of Proposition 2. We show that there exists a δ' such that for all $\delta \in (\delta', 1)$, in response to a deviation $z(\widetilde{\tau}_t)$ at t such that $\widetilde{\tau}_t < z(\widetilde{\tau}_t) < \hat{\tau}$, a punisher adopts WEC as punishment by implementing $z(\widetilde{\tau}_t)$ at $t+1$ rather than break WEC by implementing $z' > z(\widetilde{\tau}_t)$. Suppose not. Then for all $\delta \in (0,1)$ the punisher gains by instead breaking WEC, where z' is maximized at $z' = \mu(\widetilde{\tau}_t)$, given that under a suspension of GATT obligations $\hat{\tau}$ is implemented as a punishment in all future periods:

$$(1-\delta)\, w\left(\mu\left(\widetilde{\tau}_t\right), z(\widetilde{\tau}_t)\right) + \delta w\left(\hat{\tau}, \hat{\tau}\right) > (1-\delta)\, w\left(z(\widetilde{\tau}_t), z(\widetilde{\tau}_t)\right) + \delta w\left(z(\widetilde{\tau}_t), z(\widetilde{\tau}_t)\right).$$

First note that $z(\widetilde{\tau}_t) < \hat{\tau}$ and so, by A2, $w(\hat{\tau}, \hat{\tau}) < w(z(\widetilde{\tau}_t), z(\widetilde{\tau}_t))$. As the inequality holds strictly in reverse at the $\delta = 1$ limit, it must hold in reverse for a range of δ less than 1. Thus we have a contradiction. □

Proof of Proposition 3. Note that a tariff profile in which $\hat{\tau}$ is set in every period is an equilibrium path because the subgame in each period is a Nash equilibrium. But there is no incentive for either country to break a past binding by deviating upwards from such a path. So we restrict attention to a candidate equilibrium path for which $\widetilde{\tau}_t < \hat{\tau}$ at some t. Also, it may be the case that the optimal deviation from $\widetilde{\tau}_t$ entails $z(\widetilde{\tau}_t) \leq \tau_{t-1}$, in which a deviator has no incentive to break a past binding. Therefore, we only need to consider situations where $z(\widetilde{\tau}_t) > \tau_{t-1}$.

Normalize so that $t = 1$ is the first period in which $\widetilde{\tau}_t < \hat{\tau}$; let $\tau_0 = \hat{\tau}$ and $\widetilde{\tau}_1 = \widetilde{\tau} < \hat{\tau}$. We are interested in situations where a country has an incentive to break a past binding, that is, to set $\tau_t = \tau' > \tau_{t-1}$. By Proposition 2, for $\delta \in (\delta', 1)$, the punisher adheres to WEC under punishment. Suppose, contrary to the proposition, that for all $\delta \in (\delta', 1)$ a deviator gains by breaking a past binding. Then in some period $t \geq 2$ the deviator, country i, must find it optimal to set a tariff $\tau_t = \tau' > \tau_{t-1}$ given that country j sets $\tau_t = \widetilde{\tau}_t$. (Note that we use $t \geq 2$ here as there have been no *past* bindings at period $t = 1$.) If the deviator sets $\tau' > \tau_{t-1}$ then this triggers $\hat{\tau}$ in all future periods and this must yield a higher payoff than setting τ_{t-1} and facing WEC in all future periods:

$$(1-\delta)\, w\left(\tau', \widetilde{\tau}_t\right) + \delta w\left(\hat{\tau}, \hat{\tau}\right) > (1-\delta)\, w\left(\tau_{t-1}, \widetilde{\tau}_t\right) + \delta w\left(\tau_{t-1}, \tau_{t-1}\right).$$

By A1, A3, the largest possible gains from breaking a past binding occur for $\widetilde{\tau}_t = 0$. Also, $\tau_{t-1} \leq \widetilde{\tau}$. So if we can show that there exists a $\delta < 1$ for which

$$(1-\delta)\, w\left(\tau', 0\right) + \delta w\left(\hat{\tau}, \hat{\tau}\right) \leq (1-\delta)\, w\left(\tau_{t-1}, 0\right) + \delta w\left(\widetilde{\tau}, \widetilde{\tau}\right)$$

then we have established a contradiction. First note that, by A2, $w\left(\widetilde{\tau}, \widetilde{\tau}\right) \leq w\left(\tau_{t-1}, \tau_{t-1}\right)$. But $\widetilde{\tau} < \hat{\tau}$ and so, again by A2, $w\left(\hat{\tau}, \hat{\tau}\right) < w\left(\widetilde{\tau}, \widetilde{\tau}\right)$. Therefore, as the inequality holds strictly at the $\delta = 1$ limit, it must hold for a range of δ less than 1. Note that the inequality may not hold for all $\delta \in (\delta', 1)$ but only for $\delta \in (\underline{\delta}, 1)$, where $\underline{\delta} \in (\delta', 1)$. (Of course, it may be the case that $\underline{\delta} = \delta'$.) As the inequality holds for $w\left(\tau', 0\right)$, it must hold for all $\widetilde{\tau}_t > 0$ under which the gains to breaking past bindings are smaller. This establishes a contradiction. □

Proof of Lemma 3. By Lemma 2 there exists a unique solution $\overline{\tau}$ for any $\delta \in (0, 1)$. By Lemma 2(ii), $z\left(\widetilde{\tau}_t\right) \geq \overline{\tau}$ for any $\widetilde{\tau}_t \leq \overline{\tau}$ and by assumption $\overline{\tau} \geq \widetilde{\tau}_{t-1}$. Then by definition, $\chi(\widetilde{\tau}_t, \widetilde{\tau}_{t-1}) = \widetilde{\tau}_{t-1}$. □

Proof of Lemma 4. The proof is by induction. By assumption $\widetilde{\tau}_t < \widetilde{\tau}_{t-1} < \overline{\tau}$.

From (4.2) and Lemma 3, an efficient path with $\widetilde{\tau}_t \leq \overline{\tau}$, $t \geq 1$ must satisfy

$$\begin{aligned} (1-\delta)(w(\widetilde{\tau}_t, \widetilde{\tau}_t) + \delta w(\widetilde{\tau}_{t+1}, \widetilde{\tau}_{t+1}) + \ldots \quad) &= \\ (1-\delta) w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_t) + \delta w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_{t-1}),\ t &= 1, \ldots \end{aligned} \tag{A.2}$$

Advancing (A.2) one period, multiplying both sides by δ, subtracting from (A.2), and dividing the result by $1-\delta$, we get:

$$w(\widetilde{\tau}_t, \widetilde{\tau}_t) = w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_t) + \frac{\delta}{1-\delta} w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_{t-1}) - \delta \left[w(\widetilde{\tau}_t, \widetilde{\tau}_{t+1}) + \frac{\delta}{1-\delta} w(\widetilde{\tau}_t, \widetilde{\tau}_t) \right] \tag{A.3}$$

which is a second-order difference equation[39] in $\widetilde{\tau}_t$. This can be seen more clearly by rearranging (A.3) to get:

$$w(\widetilde{\tau}_t, \widetilde{\tau}_{t+1}) = \frac{1}{\delta}\left[w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_t) - w(\widetilde{\tau}_t, \widetilde{\tau}_t)\right] + \frac{w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_{t-1})}{1-\delta} - \frac{\delta w(\widetilde{\tau}_t, \widetilde{\tau}_t)}{1-\delta},\ t > 1.$$

Rewriting (4.3), we get:

$$\delta\left[w(\widetilde{\tau}_t, \widetilde{\tau}_{t+1}) - w(\widetilde{\tau}_t, \widetilde{\tau}_t)\right] = w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_t) + \frac{\delta w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_{t-1})}{1-\delta} - \left[w(\widetilde{\tau}_t, \widetilde{\tau}_t) + \frac{\delta w(\widetilde{\tau}_t, \widetilde{\tau}_t)}{1-\delta}\right]$$

By Proposition 3 and Lemma 3 we know that for $\delta \in (\underline{\delta}, 1)$ and for $\widetilde{\tau}_t < \widetilde{\tau}_{t-1} < \overline{\tau}$ the solution to the constrained optimization problem

$$\max_{\widetilde{\tau}_t < z_t < \widetilde{\tau}_{t-1}} \left\{ w(z_t, \widetilde{\tau}_t) + \frac{\delta w(z_t, z_t)}{1-\delta} \right\}$$

is $\widetilde{\tau}_{t-1}$. So we can write

$$\begin{aligned} & w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_t) + \frac{\delta w(\widetilde{\tau}_{t-1}, \widetilde{\tau}_{t-1})}{1-\delta} - \left[w(\widetilde{\tau}_t, \widetilde{\tau}_t) + \frac{\delta w(\widetilde{\tau}_t, \widetilde{\tau}_t)}{1-\delta}\right] \\ = & \max_{\widetilde{\tau}_t < z_t < \widetilde{\tau}_{t-1}} \left\{ w(z_t, \widetilde{\tau}_t) + \frac{\delta w(z_t, z_t)}{1-\delta} \right\} - \left[w(\widetilde{\tau}_t, \widetilde{\tau}_t) + \frac{\delta w(\widetilde{\tau}_t, \widetilde{\tau}_t)}{1-\delta}\right] \\ > & \ 0. \end{aligned}$$

Therefore $w(\widetilde{\tau}_t, \widetilde{\tau}_{t+1}) > w(\widetilde{\tau}_t, \widetilde{\tau}_t)$. But then, by A1, $\widetilde{\tau}_{t+1} < \widetilde{\tau}_t$ as required. □

Proof of Proposition 4. In order for $\chi(\widetilde{\tau}_t, \widetilde{\tau}_{t-1}) = \widetilde{\tau}_{t-1}$ to be the optimal deviation, by Proposition 3, we require that $\delta \in (\underline{\delta}, 1)$. Now, rewrite (4.4) as a function of $\widetilde{\tau}_1$:

$$\begin{aligned} f(\widetilde{\tau}_1) &= (1-\delta)w(\chi(\widetilde{\tau}_1, \hat{\tau}), \tau_1) + \delta w(\chi(\widetilde{\tau}_1, \hat{\tau}), \chi(\widetilde{\tau}_1, \hat{\tau})) \\ & -(1-\delta)(w(\widetilde{\tau}_1, \widetilde{\tau}_1) + \delta w(\tau_2(\hat{\tau}, \widetilde{\tau}_1), \tau_2(\hat{\tau}, \widetilde{\tau}_1)) + ...) \end{aligned}$$

Note that by the definition of $\overline{\tau}$ (see Lemma 2)

$$(1-\delta)w(\chi(\overline{\tau}, \hat{\tau}), \overline{\tau}) + \delta w(\chi(\overline{\tau}, \hat{\tau}), \chi(\overline{\tau}, \hat{\tau})) = w(\overline{\tau}, \overline{\tau})$$

Moreover, $\widetilde{\tau}_t(\hat{\tau}, \overline{\tau}) < \overline{\tau}$, all t by Lemma 4. So, if $\widetilde{\tau}_1 = \overline{\tau}$ then (4.1) is slack i.e.

$$\begin{aligned} & (1-\delta)(w(\overline{\tau}, \overline{\tau}) + \delta w(\tau_2(\hat{\tau}, \overline{\tau}), \tau_2(\hat{\tau}, \overline{\tau})) + ...) \\ > & \ w(\overline{\tau}, \overline{\tau}) = (1-\delta)w(\chi(\overline{\tau}, \hat{\tau}), \overline{\tau}) + \delta w(\chi(\overline{\tau}, \hat{\tau}), \chi(\overline{\tau}, \hat{\tau})) \end{aligned}$$

[39]This is an unusual difference equation in that it has a continuum of stationary solutions i.e. setting $\tau_{t-1} = \tau_t = \tau_{t+1}$ always solves (2.1).

where the inequality follows by A2. So, we have shown that $f(\overline{\tau}) < 0$.

Next, if $\widetilde{\tau}_1 = \varepsilon$, we have

$$(1-\delta)w(\chi(\varepsilon,\hat{\tau}),\varepsilon) + \delta w(\chi(\varepsilon,\hat{\tau}),\chi(\varepsilon,\hat{\tau})) = \max_{\varepsilon \le z \le \hat{\tau}} (1-\delta)w(z,\varepsilon) + \delta w(z,z) > w(\varepsilon,\varepsilon)$$

for ε small enough: the inequality is strict by Lemma 2 above, as for ε small enough, $z(\varepsilon) > \varepsilon$. Moreover, from Lemma 4, for ε small enough,

$$(1-\delta)(w(\varepsilon,\varepsilon) + \delta w(\tau_2(\hat{\tau},\varepsilon),\tau_2(\hat{\tau},\varepsilon)) + ...) \simeq w(\varepsilon,\varepsilon)$$

So, it is possible to choose ε small enough so that

$$(1-\delta)(w(\varepsilon,\varepsilon) + \delta w(\tau_2(\hat{\tau},\varepsilon),\tau_2(\hat{\tau},\varepsilon)) + ...) < (1-\delta)w(\chi(\varepsilon,\hat{\tau}),\varepsilon) + \delta w(\chi(\varepsilon,\hat{\tau}),\chi(\varepsilon,\hat{\tau}))$$

i.e. $f(\varepsilon) > 0$. Now, by inspection, $f(.)$ is continuous in $\widetilde{\tau}_1$ as χ and $\widetilde{\tau}_t$ are continuous in $\widetilde{\tau}_1$. So, there exists at least one value of $\widetilde{\tau}_1$ for which $f(\widetilde{\tau}_1) = 0$, and so there exists a smallest such value. □

Proof of Proposition 5. Suppose to the contrary that there exists some date $s+1$ at which $\tau_{s+1}(\hat{\tau},\tilde{\tau}_1) = 0$. By Lemma 4, the equilibrium path must be gradually decreasing. This rules out $\tau_t(\hat{\tau},\tilde{\tau}_1) = 0$ for $t > s+1$. In particular, if country i implements $\tilde{\tau}_{s+1} = 0$ with no further tariff reductions scheduled, then Lemma 4 says that country j does better by deviating in such a way that it simply fails to implement a tariff reduction at $s+1$, instead continuing to set its tariff at the level of the previous period; $\tau_{s+1} = \tilde{\tau}_s$. Thus we have a contradiction. A gradually declining tariff sequence that falls below free trade could not be efficient since, by continuity, for any such gradually decreasing tariff sequence that falls below free trade (efficiency) it is always possible to find another gradually decreasing equilibrium tariff sequence that is closer to free trade (from below).

We now prove that as $\delta \to 1$, the efficient path converges to free trade. First note that, by Lemma 2 (iii), $\lim_{\delta\to 1} \overline{\tau} = 0$. Now, by Proposition 4, at $t = 1$ $\widetilde{\tau}_1$ is chosen such that $0 < \widetilde{\tau}_1 < \overline{\tau}$. By continuity of $w(\cdot,\cdot)$, it is always possible to choose a $\widetilde{\tau}_t$ such that $0 < \widetilde{\tau}_t < \overline{\tau}$. Also by Proposition 4, on the equilibrium tariff path $\widetilde{\tau}_{t+1} < \widetilde{\tau}_t$ (without ever attaining its limit). □

A.2. Relationship of the model to the literature

The discussion that I now undertake draws on Chapter 6 of Bagwell and Staiger (2002) where enforcement of international trade agreements under GATT is discussed.[40] As explained above, Bagwell and Staiger distinguish two types of initial deviation from an agreed tariff level; on-equilibrium-path deviations and off-equilibrium-path deviations. According to their description an off-equilibrium-path deviation, familiar from the theory of repeated games, is a reversion to $\hat{\tau}$. An on-equilibrium-path deviation is a smaller deviation required to keep the incentive compatibility constraint of the agreement binding in response to an unexpected surge in import volumes. Retaliation to both types of deviation is allowed for under the dispute settlement provisions of the GATT, in particular Article XXIII. In response to a relatively small on-equilibrium-path deviation, a 'rebalancing of concessions' is allowed. This is where a trade partner simply withdraws concessions that it made to the deviator that were not reciprocated. Such a WEC can take place either under Article XIX if the measures can be agreed upon between parties themselves, or under Article XXIII if a panel is required to provide independent arbitration to help resolve a dispute in the interpretation of GATT rules.[41]

However, as for when a punishment is more severe than is sanctioned by WEC, under Article XXIII contracting parties may authorize, in 'appropriately serious cases,' a 'suspension of GATT obligations' for a deviation as well. This, Bagwell and Staiger argue, may be associated with a standard trigger-strategy punishment by reversion to $\hat{\tau}$. In their discussion, they suggest that a suspension of GATT obligations would occur only once a panel ruling had been violated, either by a deviator that raised tariffs after a panel had ruled against them, or by a country that retaliated even when the initial complaint was not upheld. However, in their formalization, a relatively large (off-equilibrium-path) deviation is met by an immediate return to trigger strategies while a relatively small (on-equilibrium-path) deviation is met by immediate WEC.

[40] A more rigorous treatment of the model in Chapter 6 is presented in Bagwell and Staiger (1990).

[41] Renegotiation of tariffs is also allowed under Article XXVIII. Under normal operation a tariff increase in one sector under Article XXVIII must be balanced by a reduction in another sector within the country, leaving the overall level of openness unchanged. But if the parties cannot agree on the renegotiation in the terms of the agreement within a specified period of time, then the party that initiated the renegotiation is free to withdraw its commitment and the other party is then limited to WEC.

Although it has an alternative formalization, the model that we develop here shares the same feature as Bagwell and Staiger's (2002 Chapter 6) in that a relatively large deviation is met by an immediate return to trigger strategies while a relatively small deviation is met by immediate WEC. The difference is that in our formalization a small deviation and retaliation by WEC is an off-equilibrium-path deviation as well i.e. once a small deviation and WEC takes place there is no return to the equilibrium path.[42]

The fact that WEC is an off-equilibrium-path deviation in the framework of this paper represents a difference in emphasis rather than a contradiction to Bagwell and Staiger's approach. Indeed, Bagwell and Staiger acknowledge that there is an off-equilibrium-path element to such a deviation because the deviator stands in violation of the agreement if it does not bring its policy back into conformity with GATT rules once the panel ruling is issued (see Bagwell and Staiger 2002, footnote 5 on p. 98). Bagwell and Staiger emphasize the written form of the GATT, particularly of Article XIX, to argue that such deviations are temporary and WEC 'legalizes' the deviation by establishing a rebalancing of concessions. Instead, I place emphasis on the observation by Dam (1970 p. 100) that "most of the tariff increases made under Article XIX have in fact never been rescinded." Dam then goes on to point out that "an affected trade partner could always demand that the concession be reinstated and may invoke the dispute settlement procedures if no action is taken." The fact that in the majority of cases no such action was taken suggests that relatively small deviations were simply matched with WEC, exactly as in the formalization of this present paper.[43]

There is a second sense in which the formalization presented here of a smaller deviation and retaliation by WEC differs from Bagwell and Staiger's. While Bagwell and Staiger back such deviation by a stochastic shock, the present modelling environment is stationary. Thus, deviation and subsequent WEC in the present framework is admittedly more difficult to justify. But again, Bagwell and Staiger identify two components to such a deviation. While their formalization emphasizes the response to the stochastic shock,

[42] Thus there are two ways that a country can trigger a suspension of GATT obligations in our model. As a deviator, it can undertake a relatively large deviation (where a 'relatively large deviation' will be defined precisely below). And as a punisher it can punish more severely than is allowed by the WEC rule (as described in Section 3).

[43] Of course, according to the theory such off-equilibrium-path deviations should never have actually been observed, but could be justified as 'trembling' or 'learning' about the way the GATT rules worked.

they also point out that an exception to an agreement under Article XIX 'raises the prospect that a government may be motivated in part by a desire to shift the costs of its intervention onto its trading partner, thus upsetting ... the balance of concessions. In its original formulation, GATT's Article XIX addresses this possibility by allowing that the trading partner can then take a retaliatory exception and withdraw its own substantially equivalent concession.' (See Bagwell and Staiger 2002 p. 105.) Thus Bagwell and Staiger clearly suggest that while part of a deviation is motivated directly by a shock, part is motivated by the desire to (opportunistically) shift costs. But presumably, providing the violation is not too big, the exception will be taken 'in good faith' and will simply be responded to by WEC. While Bagwell and Staiger formalize the part of a deviation that is in direct response to a shock, we formalize the part that is opportunistically motivated by a terms-of-trade gain (see Section 4).[44]

One aspect of the formalization may appear to go against the written procedure for dispute settlement and needs clarification. As mentioned previously, the procedures of Article XXIII center on the notion of "nullification and impairment" and do not require the actual breach of a legal obligation (i.e. breaking a scheduled or past binding.) However, as Dam (1970 p. 360) points out, dispute settlement panels came to favor "an approach that would make the legality of the trade measure under the substantive provisions of the General Agreement the crucial factor in determinations of nullification and impairment." According to this reading of events, through practice panels came to regard the breach of a legal obligation to be the key condition for nullification and impairment, as in the present formalization (Jackson 1989 backs this view; see p. 95).

A.3. An example: Quasi-linear preferences

We work through the analysis of the paper for a simple endowments model where agents have quasi-linear preferences. Each country $i \in \{1,2\}$ has an endowment (normalized to unity) of good i (or is endowed with a factor of production that can produce 1 unit of good i). We denote by x^i_j the consumption of good j in country i. The preferences of the

[44] The 'opportunistically motivated' part does not arise in Bagwell and Staiger's analysis because they analyze a symmetric situation so the terms of trade effects cancel; both countries deviate symmetrically because they both face the same shock, which both can verify. Here, by contrast, we capture an asymmetric situation in which only one country deviates, and its motives cannot be perfectly verified.

representative consumer in country i are of the following form:

$$u^i = x_i^i + \frac{\sigma}{\sigma - 1} x_j^{i\,\frac{\sigma-1}{\sigma}},\ i = 1, 2 \tag{A.4}$$

with $\sigma > 1$, and where x_j^i is consumption of good j. σ measures the elasticity if substitution between different "varieties" of imported goods. Utility is maximized subject to the budget constraint

$$p_i x_i^i + p_j (1 + \tau^i) x_j^i = p_i + R_i \tag{A.5}$$

where p_j and R_i are respectively: the world price of good j, and tariff revenue in country i which, as is usually assumed, is returned to the consumer in a lump-sum. Note that, while in the text above we referred to country i's tariff in period t as τ_t^i, here time subscripts are dropped. The optimization problem gives demands for the two goods:

$$x_j^i = \left[\frac{p_j(1+\tau^i)}{p_i}\right]^{-\sigma},\ j \neq i; \tag{A.6}$$

$$x_i^i = 1 + \frac{R_i}{p_i} - \frac{p_j(1+\tau^i)x_j^i}{p_i} = 1 + \frac{R_i}{p_i} - \left[\frac{p_j(1+\tau^i)}{p_i}\right]^{1-\sigma}, \tag{A.7}$$

where the demand for good i, x_i^i is determined residually via the budget constraint.

Indirect utility for the representative household in i is therefore derived by substituting (A.6) ,(A.7), back into (A.4) to get

$$v^i = 1 + \frac{1}{\sigma - 1}\left[\frac{p_j(1+\tau^i)}{p_i}\right]^{1-\sigma} + \frac{R_i}{p_i}. \tag{A.8}$$

Also, tariff revenue is

$$R_i = p_j \tau^i x_j^i = \frac{p_j \tau^i}{p_i}\left[\frac{p_j(1+\tau^i)}{p_i}\right]^{-\sigma}. \tag{A.9}$$

We substitute (A.9) into (A.8) to get

$$v^i = 1 + \frac{1}{\sigma - 1}\left[\frac{p_j(1+\tau^i)}{p_i}\right]^{1-\sigma} + \frac{p_j \tau^i}{p_i}\left[\frac{p_j(1+\tau^i)}{p_i}\right]^{-\sigma}. \tag{A.10}$$

We may choose p_i as the numeraire in order to write (A.10) as

$$v(\tau, p) = 1 + \frac{1}{\sigma - 1}\left[p(1+\tau)\right]^{1-\sigma} + p\tau\left[p(1+\tau)\right]^{-\sigma}. \tag{A.11}$$

Finally, we need to calculate how the (reciprocal of) terms of trade for country i, p, changes with τ', τ. Evaluating (A.6) ,(A.7) at $\tau^j = \tau'$ $\tau^i = \tau$, $p_j = p$, $p_i = 1$, we get the following:

$$x_i^i = 1 + p\tau\left[p(1+\tau)\right]^{-\sigma} - \left[p(1+\tau)\right]^{1-\sigma}; \quad \text{(A.12)}$$

$$x_i^j = \left[\frac{(1+\tau')}{p}\right]^{-\sigma}. \quad \text{(A.13)}$$

So, substituting (A.12),(A.13) into the market-clearing condition for good i, namely that supply of unity equals the sum of country demands ($1 = \sum_{i \in \{1,2\}} x_i^j$), we have

$$p\tau\left[p(1+\tau)\right]^{-\sigma} - \left[p(1+\tau)\right]^{1-\sigma} + \left[\frac{(1+\tau')}{p}\right]^{-\sigma} = 0. \quad \text{(A.14)}$$

Solving (A.14) for p, we have:

$$p(\tau, \tau') = \left(\frac{1+\tau}{1+\tau'}\right)^{\sigma/(1-2\sigma)}.$$

Note that as $\sigma > 0.5$ by assumption, $p_\tau < 0$ i.e. an increase in $i's$ tariff always improves $i's$ terms of trade. So, we may write country $i's$ indirect utility as

$$w(\tau, \tau') \equiv v(p(\tau, \tau'), \tau) = 1 + \frac{1}{\sigma - 1}\left[p(1+\tau)\right]^{1-\sigma} + p\tau\left[p(1+\tau)\right]^{-\sigma}.$$

So, a (symmetric) Nash equilibrium in tariffs is a $\hat{\tau}$ such that $v(\hat{\tau}, p(\hat{\tau}, \hat{\tau})) \geq v(\tau, p(\tau, \hat{\tau}))$, all $\tau \neq \hat{\tau}$.

As v is continuously differentiable, we can characterize $\hat{\tau}$ as the solution to

$$v_\tau(\hat{\tau}, p(\hat{\tau}, \hat{\tau})) + v_p(\hat{\tau}, p(\hat{\tau}, \hat{\tau}))p_\tau(\hat{\tau}, \hat{\tau}) = 0, \quad \text{(A.15)}$$

where v_τ, v_p denote partial derivatives of v. Now,

$$v_\tau(\tau, p) = -\sigma\tau p^{1-\sigma}(1+\tau)^{-\sigma-1}; \quad \text{(A.16)}$$

$$v_p(\tau, p) = -p^{-\sigma}(1+\tau)^{1-\sigma} + (1-\sigma)p^{-\sigma}\tau(1+\tau)^{-\sigma};$$

$$p_\tau = \frac{\sigma}{1-2\sigma}\left(\frac{1+\tau}{1+\tau'}\right)^{(\sigma/(1-2\sigma))-1}\frac{1}{1+\tau'}.$$

So, using (A.16) and the fact that $p(\hat{\tau}, \hat{\tau}) = 1$, we have from (A.15) that

$$-\sigma\hat{\tau}(1+\hat{\tau})^{-\sigma-1} + \left[-(1+\hat{\tau})^{1-\sigma} + (1-\sigma)\hat{\tau}(1+\hat{\tau})^{-\sigma}\right]\frac{\sigma}{1-2\sigma}\frac{1}{1+\hat{\tau}} = 0.$$

Eliminating common terms, we get

$$-\hat{\tau}+[-(1+\hat{\tau})+(1-\sigma)\hat{\tau}]\frac{1}{1-2\sigma}=0.$$

Solving for $\hat{\tau}$, we get

$$\hat{\tau}=\frac{1}{\sigma-1}$$

for the optimal tariff. Recall that $\sigma>1$, so $\hat{\tau}$ is defined and positive.

Now we have $\hat{\tau}$, we can check that A1, A2 and A3 hold.

Substituting for $p(\tau,\tau')$, we can write the payoff function as follows:

$$w(\tau,\tau')=1+\left(\frac{(1+\tau)^{1-\sigma}}{\sigma-1}+\tau(1+\tau)^{-\sigma}\right)\left(\frac{1+\tau}{1+\tau'}\right)^{\sigma(1-\sigma)/(1-2\sigma)}.$$

We can use this expression to verify that A1 and A2 hold. Note that A3 is not satisfied globally but is satisfied for $\tau\in[0,\hat{\tau})$, which is sufficient for existence of equilibrium. Uniqueness must then be checked "by hand" but is confirmed by the uniqueness of the solution for $\hat{\tau}$ presented above.

Take A1 first:

$$w_1(\tau,\tau')=\frac{\sigma\left(1+\tau\right)^{-1-\sigma}\left(1-\left(\sigma-1\right)\tau\right)}{2\sigma-1}\left(\frac{1+\tau}{1+\tau'}\right)^{\sigma(1-\sigma)/(1-2\sigma)}.$$

The sign of this expression depends on the term in brackets $(1-(\sigma-1)\,\tau)$. If $\tau=\hat{\tau}=1/\left(\sigma-1\right)$ and so $(1-(\sigma-1)\,\tau)=0$ then $w_1(\tau,\tau')=0$. If $\tau<\hat{\tau}$ then $(1-(\sigma-1)\,\tau)>0$ and so $w_1(\tau,\tau')>0$, while if $\tau>\hat{\tau}$ then $(1-(\sigma-1)\,\tau)<0$ as required.

$$w_2(\tau,\tau')=-\frac{\sigma\left(1+\tau\right)^{-1-\sigma}\left(1+\sigma\tau\right)}{2\sigma-1}\left(\frac{1+\tau}{1+\tau'}\right)^{\left(1-\sigma-\sigma^2\right)/(1-2\sigma)}<0\text{ for all }\tau,\tau'>0.$$

Now A2:

$$w_1(\tau,\tau')+w_2(\tau,\tau')=$$
$$-\frac{\sigma\left(1+\tau\right)^{-2-\sigma}\left(\sigma\tau\left(2+\tau+\tau'\right)-\left(1+\tau\right)\tau'\right)}{2\sigma-1}\left(\frac{1+\tau}{1+\tau'}\right)^{\left(1-\sigma-\sigma^2\right)/(1-2\sigma)}$$

Setting $\tau'=\tau$ as in A2, we get

$$w_1(\tau,\tau)+w_2(\tau,\tau)=-\sigma\tau\left(1+\tau\right)^{-1-\sigma}.$$

It is easy to see that when $\tau = \tau' = 0$ we have $w_1(0,0) + w_2(0,0) = 0$. This is necessary for free trade to maximize efficiency. Moreover, $w_1(\tau,\tau) + w_2(\tau,\tau) < 0$ for all $\tau > 0$, $\sigma > 1$, as required.

Finally, we show that a slightly weaker version of A3 holds; $w_{12}(\tau,\tau') < 0$ for $\tau \in (0,\hat{\tau})$ and that although $w_{12}(\tau,\tau') > 0$ for $\tau > \hat{\tau}$, the uniqueness of the solution for $\hat{\tau}$ tells us that the

$$w_{12}(\tau,\tau') = -\frac{(\sigma-1)\,\sigma^2\,(1+\tau)^{-2-\sigma}\,(1-(\sigma-1)\,\tau)}{(2\sigma-1)^2}\left(\frac{1+\tau}{1+\tau'}\right)^{\sigma(1-\sigma)/(1-2\sigma)}.$$

So $w_{12}(\tau,\tau') < 0$ for $\tau \in [0,\hat{\tau})$ because $(1-(\sigma-1)\,\tau) > 0$.

Now we want to characterize the constrained deviation, using it to derive $\overline{\tau}$. Setting this first order condition equal to zero, we have

$$w_1(z(\tau),\tau) + \frac{\delta}{1-\delta}(w_1(z(\tau),z(\tau)) + w_2(z(\tau),z(\tau))) = 0.$$

We can write (A.4) as follows

$$w(z(\tau),\tau) = 1 + \left(\frac{1+z(\tau)}{1+\tau}\right)^{\sigma(1-\sigma)/(1-2\sigma)}\gamma(z(\tau)),$$

where $\gamma(z(\tau)) = \frac{(1+z(\tau))^{1-\sigma}}{\sigma-1} + z(\tau)\,(1+z(\tau))^{-\sigma}$, so $\gamma'(z(\tau)) = -\sigma z(\tau)\,(1+z(\tau))^{-1-\sigma}$ .Then

$$w_1(z(\tau),\tau) = \frac{\frac{\sigma(1-\sigma)}{1-2\sigma}w(z(\tau),\tau)}{(1+z(\tau))} + \left(\frac{1+z(\tau)}{1+\tau}\right)^{\frac{\sigma(1-\sigma)}{1-2\sigma}}\gamma'(z(\tau)),$$

and

$$w_2(z(\tau),\tau) = -\frac{\frac{\sigma(1-\sigma)}{1-2\sigma}w(z(\tau),\tau)}{(1+\tau)}.$$

It is then straightforward to see that the first order condition can be rewritten $(1-\delta)\,w_1(z(\tau),\tau) + \delta\gamma'(z(\tau)) = 0$. Setting $z(\tau) = \tau = \overline{\tau}$ in the first order condition, we get

$$(1-\delta)\,\frac{\sigma(\sigma-1)}{2\sigma-1}\frac{\gamma(\overline{\tau})}{1+\tau^*} + \gamma'(\overline{\tau}) = 0.$$

Substituting for $\gamma(\overline{\tau})$ and $\gamma'(\overline{\tau})$ and simplifying, the equation becomes

$$\frac{\sigma\,(1+\overline{\tau})^{-1-\sigma}\,(1-\delta+(1-\sigma\,(1+\delta))\overline{\tau})}{2\sigma-1} = 0.$$

Solving, the only admissible root[45] is

$$\overline{\tau} = \frac{1-\delta}{\sigma\,(1+\delta)-1}.$$

[45] The root $\tau = -1$ also solves this expression.

References

[1] Bagwell, K., Staiger, R., 1990. A theory of managed trade. American Economic Review 80, 779-795.

[2] Bagwell, K., Staiger, R., 1996. Reciprocal trade liberalization. NBER Working Paper No. 5488.

[3] Bagwell, K., Staiger, R., 1999. An economic theory of the GATT. American Economic Review 89, 215-48.

[4] Bagwell, K., Staiger, R., 2001. Reciprocity, non-discrimination, and preferential agreements in the multilateral trading system. European Journal of Political Economy 17, 281-325.

[5] Bagwell, K., Staiger, R., 2002. The Economics of the World Trading System. MIT Press, Cambridge, Massachusetts.

[6] Bhagwati, J., 1998. Protectionism. MIT Press, Cambridge, Massachusetts.

[7] Blume, A., 1994. Intraplay communication in repeated games. Games and Economic Behavior 6, 181-211.

[8] Bond, E.B., Park, J.H., 2002. Gradualism in trade agreements with asymmetric countries. Review of Economic Studies 69, 379-406.

[9] Bond, E.B., Riezman, R., and Syropoulos, C., 2004. A strategic and welfare theoretic analysis of free trade areas. Journal of International Economics 64, 1-27.

[10] Bond, E.B., Syropoulos, C., Winters, A., 2001. Deepening of regional integration and multilateral trade agreements. Journal of International Economics 53, 335-361.

[11] Chisik, R., 2003. Gradualism in free trade agreements: A theoretical justification. Journal of International Economics 59, 367-397.

[12] Conconi, P., Perroni C., 2004. The Economics of Special and Differential Trade Regimes. CEPR Discussion Paper no. 4508.

[13] Dam, K.W., 1970. The GATT: Law and International Economic Organization. Chicago University Press, Chicago.

[14] Devereux, M., 1997. Growth, specialization and trade liberalization. International Economic Review, 38, 565-585.

[15] Dixit, A., 1987. Strategic Aspects of Trade Policy, in: Bewley, T.F. (Ed), Advances in Economic Theory; Fifth World Congress. Cambridge University Press, Cambridge.

[16] Farrell, J., Maskin, E., 1989. Renegotiation in repeated games. Games and Economic Behavior 1, 327-360.

[17] Friedman, J., 1971. A noncooperative equilibrium for supergames. Review of Economic Studies 38, 1-12.

[18] Fudenberg, D., Tirole, J., 1991. Game Theory. MIT Press, Cambridge Massachusetts.

[19] Furusawa, T., and Lai, E.L.-C., 1999. Adjustment costs and gradual trade liberalization. Journal of International Economics 49, 333-361.

[20] GATT, 1994. Uruguay Round: Final act. Marrakesh, 15 April. Photocopy.

[21] Grossman, G., Helpman, E., 1995. Trade wars and trade talks. Journal of Political Economy 103, 675-708.

[22] Hamilton, R., Whalley, J., 1983. Optimal tariff calculations in alternative trade models and possible implications for current world trading arrangements. Journal of International Economics 15, 323-348.

[23] Jackson, J., 1969. World Trade and the Law of the GATT; (a legal analysis of the General Agreement on Tariffs and Trade). Bobbs-Merril, Indianapolis.

[24] Jackson, J., 1989. The World Trading System. MIT Press, Cambridge Massachusetts.

[25] Johnson, H.G., 1953-54. Optimum tariffs and retaliation. Review of Economic Studies 21, 142-153.

[26] Lockwood, B., Thomas, J., 2002. Gradualism and irreversibility. Review of Economic Studies 69, 339-356.

[27] Lockwood, B., Wong, K.Y., 2000. Specific and ad valorem tariffs are not equivalent in trade wars. Journal of International Economics 52, 183-195.

[28] Lockwood, B., Zissimos, B., 2002. Trade agreements with limited punishments. Centre for the Study of Globalisation and Regionalisation Discussion Paper 87/02, University of Warwick.

[29] Maggi, G., Rodriguez-Clare, A., 2005. A political-economy theory of trade agreements. NBER Working Paper no. 11716.

[30] Mayer, W., 1981. Theoretical considerations of negotiated tariff adjustments. Oxford Economic Papers 33, 135-153.

[31] McCutcheon, B., 1997. Do meetings in smoke-filled rooms facilitate collusion? Journal of Political Economy 105, 330-350.

[32] McLaren, J., 1997. Size, sunk costs and Judge Bowker's objection to free trade. American Economic Review 87, 400-420.

[33] Scherer, F.M., 1980. Industrial Market Structure and Economic Performance. Houghton Mifflin.

[34] Staiger, R., 1995a. International rules and institutions for trade policy, in: Grossman, G., Rogoff, K., (Eds), Handbook of International Economics, Volume III, North Holland.

[35] Staiger, R., 1995b. A theory of gradual trade liberalization, in: Levinsohn, J., Deardorff, A.V., Stern, R.M. (Eds.), New Directions in Trade Theory. University of Michigan Press, Ann Arbor, MI.

[36] Syropoulos, C., 2002. Optimum tariffs and retaliation revisited: How country size matters. Review of Economic Studies 69, 707-727.

[37] van Damme, E., 1989. Renegotiation-proof equilibria in repeated prisoners' dilemma. Journal of Economic Theory 47, 206-217.

[38] Whalley, J., Hamilton, C., 1996. The Trading System After The Uruguay Round. Institute for International Economics, Washington, DC.

Figure 1

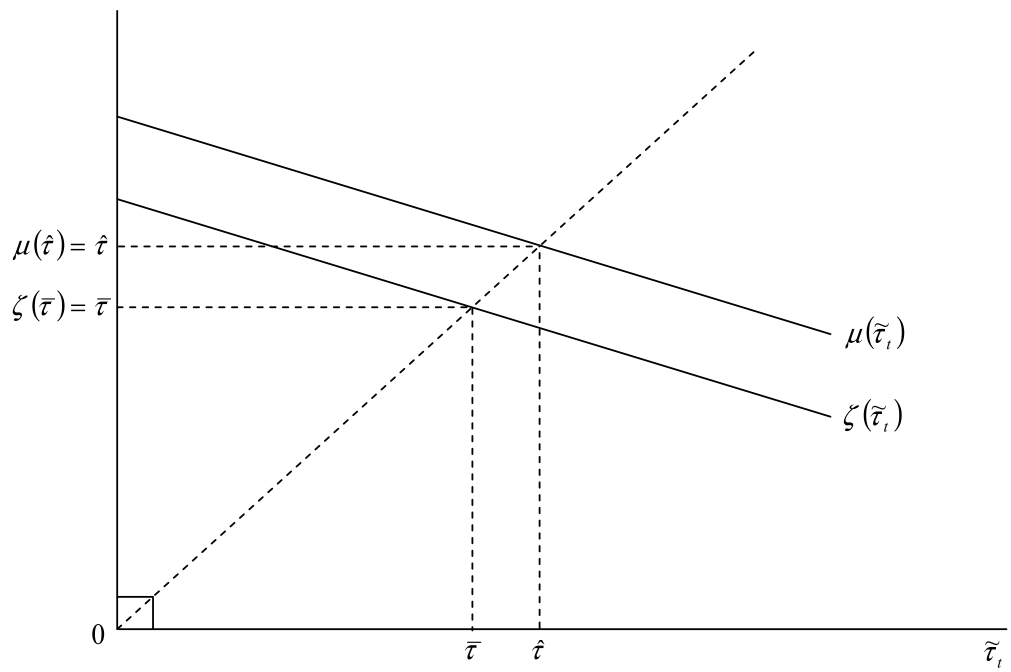

Figure 2

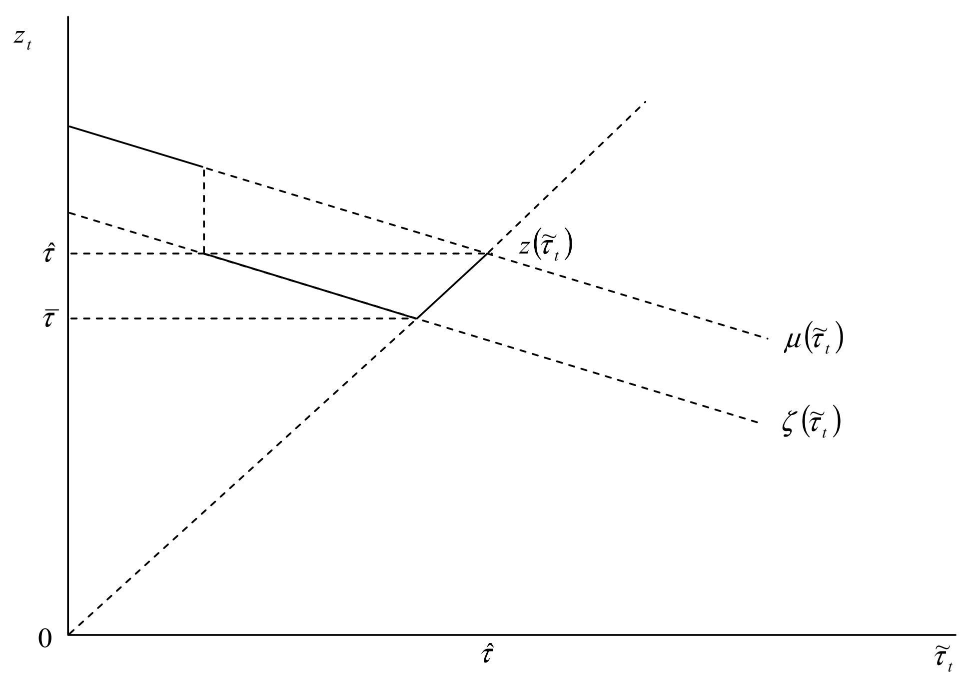

Figure 3

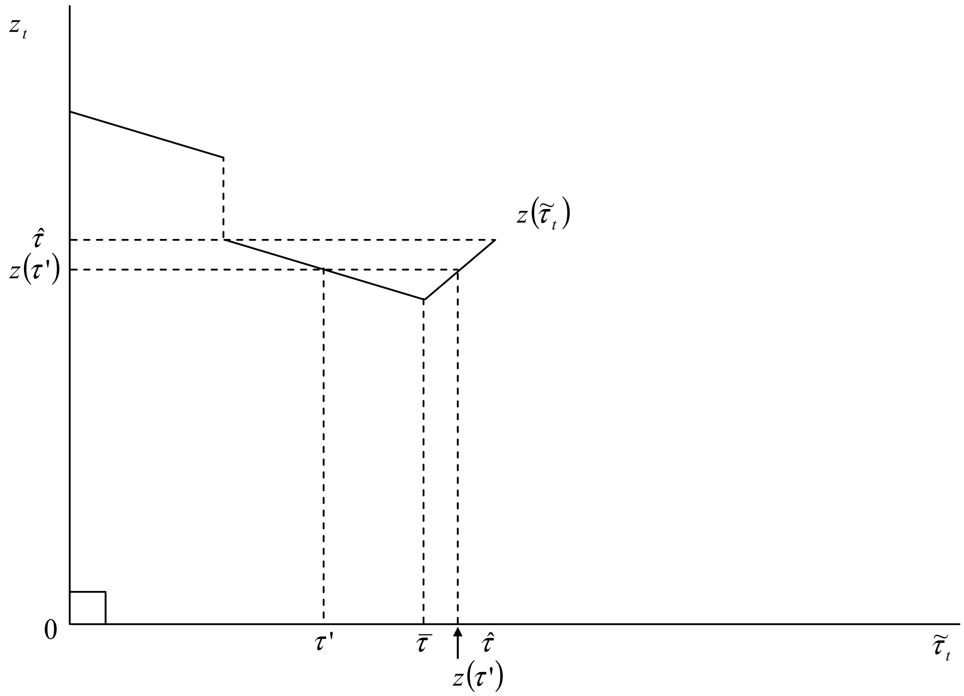

Figure 4