

Optimal Time Limits on Safeguards in Trade Agreements

Mathias Herzing^{*†}

Abstract

This paper addresses the issue of having time limits on how long countries should be permitted to withdraw liberalization commitments under a trade agreement. In a setting with two countries and a continuum of sectors, each sector is subject to stochastic switches between two states over time. Under trade liberalization, there are gains to be made in the good state, while losses will be incurred when being in the bad state and protection by means of a safeguard is thus desirable. It is shown that, by limiting the time the safeguard can be applied, the interests of winners and losers in liberalization are balanced across countries. However, an ex ante agreed-upon finite time limit on the use of the safeguard will eventually be perceived as too short, as the share of sectors in need of being exempted from it increases over time.

^{*}Institute for International Economic Studies, Stockholm University, August 2005

[†]I am grateful to my advisors Harry Flam and Henrik Horn. Helpful comments from Thomas Eisensee and Per Krusell, as well as from participants in the Nordic International Trade Seminar in Helsinki in May 2005 are appreciated. Thanks also to Christina Lönnblad for editorial assistance. Financial support from Jan Wallander's and Tom Hedelius' Research Foundation is gratefully acknowledged.

1 Introduction

International agreements in general, and trade agreements in particular, typically include safeguard provisions, which allow a country to (usually temporarily) withdraw a concession made under the agreement in certain contingencies. More specifically, a safeguard makes it possible for a country to introduce protectionist measures, thereby scaling back the agreed-upon liberalization.

This paper addresses the issue of prespecifying the length of the phase, in which a country is permitted to apply a safeguard. The focus is on two issues. First, an analytical framework for how time limits on safeguards are determined is provided and second, the ex post implications of a time limit on the use of safeguards within this framework are examined.

In the present model, there are two symmetric countries with a continuum of sectors, each of which can be in either of two states. In the good state, there are gains from trade liberalization to be made, while in the bad state, losses will be incurred under trade liberalization. It is in the latter case that invoking a safeguard, allowing for scaling back liberalization, is desirable. Shifts from one state to the other are assumed to be solely stochastically determined by Poisson processes that are identical and independent across countries and between sectors. Hence, there is no upper limit to how long a country can remain in the bad (or the good) state.

Negotiations over a trade agreement between the countries are assumed to cover the provision of a safeguard to be applied during a phase, henceforth referred to as the adjustment phase. The determination of cooperative tariff levels will not be addressed. Hence, the analysis will focus on the optimal choice of length of the adjustment phase for any given cooperative tariffs.

Ex ante, when the agreement is negotiated, a country must weigh two effects of a safeguard against each other. On the one hand, it may wish to implement the safeguard for as long as it deems necessary to stem losses from liberalization. On the other hand, it may find itself exposed to the safeguard measures of another country, in which case it will prefer the safeguard to be applied for as short a period as possible to avoid foregoing the gains from liberalization.

In the negotiations, the two countries are assumed to agree on a rule prespecifying the maximum duration for applying the safeguard. It is

shown that what solely determines the optimal length of the adjustment phase is the ratio between the gain from liberalization in the good state and the loss from liberalization in the bad state. If this ratio is smaller than or equal to one, it is optimal to let the adjustment phase be infinite. If it takes on values between one and a threshold value depending on the rate of discounting, letting the length of the adjustment phase be strictly positive, but finite is optimal; the optimal length of the adjustment phase decreases monotonously in this interval of ratios. If the ratio between the gain from liberalization in the good state and the loss from liberalization in the bad state exceeds the threshold value, it is optimal not to allow for an adjustment phase at all, i.e. not to include a safeguard in the agreement. The intuition is straightforward. The larger are the gains from liberalization vis-à-vis the losses, the stronger is the incentive to limit the length of the adjustment phase ex ante.

In the case of a finite adjustment phase, it is shown that the discounted value of future average sector payoffs will eventually start falling. In fact, the time limit on the use of the safeguard will become suboptimal, and its ex post globally optimal value will increase over time. The intuition behind this result is that ex ante, too low a weight is attributed to the situation when a sector has been in the bad state longer than the agreed-upon adjustment phase and hence, would benefit from further protection. The underlying reason is that this situation can only emerge after the agreement has lasted longer than the adjustment phase. In fact, the likelihood of being in that situation increases over time, once the agreement has lasted longer than the adjustment phase. As the likelihood of being in need of extended protection is increasing, the agreed-upon adjustment phase will therefore eventually be too low in relation to what is globally optimal, and the discounted value of future average sector payoffs will be lower than what it would be under the ex post optimal adjustment phase length.

A politically interesting implication of ex post suboptimality is that, in the absence of any readjustment possibilities, the dissatisfaction with the agreement will increase over time, as there is an increase in the share of sectors in the bad state for a period longer than the adjustment phase. What once seemed optimal will, to an increasing degree, be viewed as inappropriate. The pressure to resort to other means of protection might increase, as the share of sectors being in the bad state for a longer period than the adjustment phase increases. Hence, the use of extralegal forms of protection might increase over time. However, while addressing the potential implications of the results derived from the model presented here, resolving them is beyond the scope of the present model.

Whereas hitherto most contributions to the role of safeguards have assumed strategic interaction to take place in infinitely repeated Prisoner's Dilemma settings¹, the present paper will apply a different methodological approach. The underlying assumption of infinitely repeated Prisoner's Dilemma games is that retaliation against deviations are, by necessity, always delayed. Such an assumption, implying that it is possible to make short-term gains by deviating against trading partners, is problematic. Criticizing this approach, Ethier (2001) emphasizes that contemporary technology and politics should actually make it possible to instantly punish deviation, thus eliminating the opportunities for short-term gains.²

In the present setting, short-term gains are not possible and hence, any deviation can instantly be punished. A government contemplating deviation must thus weigh the immediate response by its trading partner into its decision. If it chooses to deviate, it will do so because it is better off under deviation-cum-retaliation than under mutual cooperation. In the present model, it will be assumed that such situations, under which the strategic interaction is no longer of Prisoner's Dilemma type, may emerge. Hence, while in the aforementioned contributions the role of safeguards is to counter the incentives to deviate in an infinitely repeated Prisoner's Dilemma setting, safeguards are included to alleviate damages from liberalization and time limits on their application serve to strike a balance between winners and losers from liberalization in the present framework.

The next section provides a background to the present paper. Section 3 describes the model with two countries and a continuum of sectors. In section 4, trade liberalization is introduced into this setting, and optimality conditions are derived. In the following section, the ex post implications of a time limit on the use of safeguards are addressed. Section 6 concludes.

2 Safeguards in Trade Agreements

A safeguard under a trade agreement is a provision allowing a signatory member to withdraw or cease to apply its normal obligations in order to protect certain overriding interests under specified conditions.

¹See, for example, Rosendorff and Milner (2001), Herzing (2005b), Hochman (2004) and Martin and Vergote (2004).

²While it may be unrealistic to assume that trade agreements serve to solve a Prisoner's Dilemma problem, it may still be the case that a trade agreement introduces Prisoner's Dilemma type of interaction by prescribing reactions to be delayed.

Here, the focus will be on the safeguard provisions under the General Agreement on Tariffs and Trade (GATT), with particular emphasis on what is usually referred to as the “escape clause”, Article XIX of the GATT. Article XIX of the GATT specifically addresses situations where a country suffers from sudden import surges seriously threatening domestic industries and which may thus be exposed to the temptation to break commitments made under the GATT (see World Trade Organization (1994a)). To avoid deviations from the agreement in such situations, Article XIX §1(a) provides the possibility of temporarily suspending obligations under the agreement to prevent or remedy injury due to liberalization commitments.

Besides the establishment of the World Trade Organization (WTO), the Uruguay Round (1986-1994), among other things, resulted in the Agreement on Safeguards, which contains rules governing the use of safeguard measures, specifically those pursuant to Article XIX (see World Trade Organization (1994b)). The Agreement on Safeguards prescribes safeguard measures to generally be on a Most-Favored-Nation (MFN) basis, although selective applications are permitted (Article 5:2b). Clearly defined time limits on the use of safeguard measures are also specified. Safeguard measures are only permitted for a period not exceeding four years (Art 7:1), except under special circumstances (Art 7:2); the total period of application shall not exceed eight years, however (Art 7:3).³ Having applied a safeguard measure, it cannot be reinvoked “for a period of time equal to that during which such measure had been previously applied, provided that the period of non-application is at least two years” (Art 7:5). However, an exception is made if the safeguard measure has a duration of 180 days or less, at least one year has elapsed since the introduction of the initial measure and such a measure has not been applied more than twice during five years preceding its date of introduction (Art 7:6).

The starting point for the analysis of the inclusion of safeguards in international agreements on cooperation between countries is the observation that there may be ex post incentives to make adjustments to the commitments made under such an agreement. Such incentives may arise due to unforeseen events making the outcome under the ex ante agreed-upon cooperative regime suboptimal from a single country’s point-of-view. In the context of a trade agreement, the ex ante negotiated degree of liberalization may turn out to be ex post suboptimal for an individual

³Developing countries are granted the possibility of extending a safeguard measure up to ten years, inclusive of extensions (Art 9:2).

country. In the absence of any adjustment instrument, the incentives to breach the agreement may become sufficiently large to make the cooperative regime unsustainable. The inclusion of a safeguard (or any other flexibility-enhancing instrument) can therefore be justified by ex ante uncertainty about what contingencies may arise ex post.⁴

3 The Model

There are two countries with a continuum of symmetric sectors, where each sector is engaged in trade with the corresponding sector in the other country. The assumption of a continuum of sectors is made for analytical convenience. Due to symmetry, it suffices to focus on one country only.

Time is taken to be continuous.⁵ Let the instant payoff of the home country government in each sector $i \in [0, 1]$ be denoted by w^i . Each sector i is assumed to be in either of two states, i.e. $\varepsilon^i \in \{\underline{\varepsilon}, \bar{\varepsilon}\}$, where ε^i is stochastically determined and perfectly observable. Let t^i and t^{i*} be the home and the foreign country's tariffs in sector i , respectively. For tractability, the following assumption will be made.

Assumption 1 w^i is independent of all stochastic variables, except ε^i , for any given pair of sector tariffs (t^i, t^{i*}) .

This simplifying assumption has several strong implications. First, it implies that there are no externalities across sectors. With a continuum of sectors, the overall impact of other sectors can reasonably be assumed to be neutral, however. Second, it implies that the state of the corresponding sector in the other country has no impact on domestic sector payoffs. Allowing for positive (or negative) externalities will not qualitatively alter the results obtained.⁶

Let the payoff generated in sector i be defined as $w^i \equiv w^i(\varepsilon^i, t^i, t^{i*})$. For any pair of t^i and t^{i*} , it is the case that $w^i(\bar{\varepsilon}, t^i, t^{i*}) > w^i(\underline{\varepsilon}, t^i, t^{i*})$.

⁴Several contributions have introduced various types of uncertainty into models of strategic interaction between trading countries, e.g. Feenstra (1987), Bagwell and Staiger (1990), Jensen and Thursby (1990), Feenstra and Lewis (1991), Riezman (1991) and Herzing (2005a).

⁵The use of continuous rather than discrete time is convenient when calculating the length of the optimal adjustment phase, although there is no qualitative difference between the results obtained under these two different approaches.

⁶Assuming the number of countries to be very large, in which case the aggregate impact of foreign realizations of the stochastic variable would be constant could, for example, justify assumption 1. A large number of countries would, however, complicate the analysis in other respects.

Thus, the realization of ε^i can, for example, be regarded as reflecting high/low domestic demand or high/low productivity. Henceforth, $\bar{\varepsilon}$ ($\underline{\varepsilon}$) will be referred to as the good (bad) state.

Switches between states (in both directions) are assumed to be governed by Poisson processes that are identical and independent between sectors and across countries. Hence, whenever a sector is in one state, the likelihood p of a switch of states at time T in the future is given by $\rho(T) = 1 - e^{-T}$. The likelihood of a switch is thus independent of how long the country has already been in a state. Theoretically, it is possible that a sector will remain in a state for any finite length of time.

When the game starts, half of the sectors are assumed to be in the bad state⁷, while it is random which sectors are actually in that state. Then, each sector will find itself in either state for varying lengths of time, depending on the stochastic process. a continuum of sectors implies that half the sectors will be in the good (bad) state at any point in time. Thus, both countries will always be equally well off, and will never have diverging interests.

In the absence of a trade agreement, both countries will in each sector apply their optimal tariff t_N vis-à-vis each other which, for simplicity, is assumed to be trade-excluding. The home country's instant payoff from sector i in the absence of any trade cooperation is given by $w_N(\underline{\varepsilon})$ or $w_N(\bar{\varepsilon})$, depending on in which state it is. The assumption of t_N being trade-excluding is made to avoid that t_N is state-dependent and hence, that externalities across countries in corresponding sectors arise.⁸

Let τ be the time since the agreement was implemented, and let $\bar{w}(\tau)$ be the average instant sector payoff at time τ .⁹ In the absence of trade cooperation, it is thus the case that

$$\bar{w}_N(\tau) = \frac{1}{2}[w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon})].$$

The common rate of discounting future payoffs is given by $\delta \in (0, 1)$, thereby implying a discount factor of $1 - \delta$. Let $\bar{v}(\tau)$ be the average

⁷Since the likelihood of switching states is the same in either state, this assumption is necessary to keep the shares of sectors in the good (bad) state constant over time.

⁸If t_N and t_N^* were instead state-dependent, i.e. $t_N = t_N(\varepsilon)$ and $t_N^* = t_N^*(\varepsilon^*)$, the Nash payoffs would not only depend on the domestic state, but also indirectly, through the foreign tariff, on the state of the trading partner. Having a continuum of sectors, however, $w_N(\underline{\varepsilon})$ and $w_N(\bar{\varepsilon})$ could in this case be interpreted as the average payoffs in the respective state. Or, alternatively, $w_N(\underline{\varepsilon})$ and $w_N(\bar{\varepsilon})$ could be seen as expected values.

⁹With a finite number of sectors, this would be the expected payoff at time τ .

discounted flow of sector payoffs at time τ , which in the absence of any trade agreement is given by

$$\bar{v}_N(\tau) = \frac{1}{\delta} \bar{w}_N(\tau) = \frac{1}{2\delta} [w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon})].$$

The two countries may agree to lower trade barriers in all sectors, i.e. to agree upon a tariff $t_C < t_N$. Let $w_C(\bar{\varepsilon})$ and $w_C(\underline{\varepsilon})$ be the payoffs generated by a sector in states $\bar{\varepsilon}$ and $\underline{\varepsilon}$, respectively, under a commonly agreed-upon cooperative tariff. The following assumption is crucial.

Assumption 2 $w_C(\bar{\varepsilon}) > w_N(\bar{\varepsilon})$
 $w_C(\underline{\varepsilon}) < w_N(\underline{\varepsilon})$.

This assumption implies that a sector will be better off under liberalization only if it finds itself in state $\bar{\varepsilon}$; being in state $\underline{\varepsilon}$, it will actually be worse off than in the absence of liberalization. This assumption can, for example, be justified if the realization of state $\underline{\varepsilon}$ corresponds to the firms in the sector being uncompetitive, while state $\bar{\varepsilon}$ corresponds to their being competitive. In the former case, a country's sector will lose from liberalization while, in the latter case, it will benefit from lower trade barriers.

It will be assumed that there are no short-term gains to be made, in accordance with, for example, Ethier (2002). In other words, any deviation will be followed by instant retaliation. Hence, interaction is not of Prisoner's Dilemma type.

Since a sector will be worse off under liberalization, it may, ex ante, when an agreement on liberalization is negotiated, be the case that both countries wish to include a safeguard, allowing for scaling back liberalization in a bad-state sector for a period of time of length λ . On the one hand, a government will want the safeguard to be applied as long as necessary, i.e. until a bad-state sector switches to state $\bar{\varepsilon}$, in which case it will prefer the agreed-upon degree of liberalization for that sector. On the other hand, a government will wish λ to be low, in case a sector is in state $\bar{\varepsilon}$, because good-state sectors unambiguously gain from liberalization.¹⁰

It must be emphasized that agreements with a predefined length of the adjustment phase constitute a subclass of all possible agreements.

¹⁰This is reminiscent of the reciprocal-conflict problem, addressed by Ethier (2001), which arises because of conflicting interests with regard to the degrees of punishment to be allowed under a trade agreement.

Within this subclass, there is a clear-cut rule prescribing exactly for how long a safeguard can be applied, whenever a switch to the bad state has occurred. Applying the safeguard under other conditions, i.e. after the adjustment phase has elapsed, is thus explicitly forbidden and regarded as a breach of the agreement. Alternatively, agreements with more flexible rules governing the use of safeguards could be considered.¹¹ However, there are strong reasons to assume that the approach taken here, with strictly defined time limits for the use of safeguards under any circumstances, is relevant. After all, the GATT-WTO is a rules-based concept. In fact, as pointed out in the previous section, there are explicit time limits for the use of safeguards under the GATT-WTO. But also from a more practical-realistic point of view, a clear-cut rule, to which participants are required to adhere, may be preferable to a system allowing for more flexibility. In the latter case, there is a danger either that the length of application of the safeguard would be left to countries' discretion, or that coordination across countries would be required (something that might be associated with some cost or be politically unfeasible). While addressing the potential suboptimality of the derived optimal solution under a clear-cut rule, the present paper abstracts away from other subclasses of safeguard agreements.

For tractability, the following assumption will be made.

Assumption 3 Whenever the safeguard is invoked by one country, both countries revert to the Nash equilibrium in that sector.

There are two important aspects of this assumption. First, it implies that applying the safeguard is equivalent to entirely scaling back liberalization. Naturally, this is a strong assumption. In reality, by imposing protection under a safeguard, countries only revert liberalization to some extent, depending on how large an injury an industry is perceived to suffer. But since it is assumed in the present model that the injury in the bad state exactly equals the concession made under liberalization, assuming a complete withdrawal of liberalization is in accordance with Article XIX §1(a) of the GATT (see section 2).

Furthermore, assumption 3 implies that scaling back liberalization will be mutual rather than unilateral. Hence, a country scaling back liberalization in one sector will also face such a scale-back in the corresponding sector of the other country. Alternatively, and possibly more realistically with regard to the legal provisions governing the escape from

¹¹The optimal solution in the subclass under consideration may actually not be first-best. This will be more thoroughly discussed in the next section.

commitments under the GATT (see World Trade Organization (1994a)), it could be assumed that the realization of $\underline{\varepsilon}$ leads to unilateral deviation for a maximum duration of λ . Making the assumption that both countries will actually deviate does not qualitatively alter the analysis, however.^{12,13}

Thus, it will be the case that, once one sector finds itself in state $\underline{\varepsilon}$, both countries will entirely scale back liberalization for a maximum duration of λ . After this time interval has elapsed, or if state $\bar{\varepsilon}$ is reached before that, both countries will revert to the agreed-upon degree of liberalization in that sector.

It is implicitly assumed that a safeguard can be applied as soon as a sector switches to state $\underline{\varepsilon}$, irrespective of how much time has elapsed since it was previously used. As pointed out in section 2, there are clearly defined limits for when a safeguard measure can be reinvoked. For tractability, this legal restriction on the possibility to apply a safeguard will not be considered here.

4 Optimal Time Limits on Protection under Trade Liberalization

4.1 Implementing a Safeguard

As emphasized in the previous section, a continuum of sectors implies that both countries will always have identical aggregate payoffs, although sector payoffs may differ within and between the two countries. Hence, the winners and losers from liberalization will be sectors within a country rather than entire countries, and conflicting interests may therefore arise within rather than between countries.

Once the agreement is in place, the average sector payoff will depend on how large a share of the sectors actually apply the safeguard. Since

¹²Under the present assumption of perfect observability, the legitimate use of the safeguard is guaranteed, because any attempt to make a gain by applying the safeguard in the good state can be punished. In the presence of hidden information, however, assumption 3 could be justified as a means of providing the correct incentives for using the escape clause.

¹³What is of importance is that the safeguard-applying country benefits, while its trading partner loses as long as the safeguard is implemented. Assuming that both countries in fact deviate when a safeguard is applied has the same effect on payoffs, although not on the size of these effects. The safeguard-applying country gains, but to a lower degree than under unilateral deviation, while the trading partner loses, albeit less so than under unilateral deviation.

there is a continuum of sectors, half of all sectors will be in the bad state at any point in time, but not all the bad-state sectors will eventually be allowed to apply the safeguard if $\lambda < \infty$. Those sectors that have been in state $\underline{\varepsilon}$ for a period of time exceeding λ will no longer be protected and hence, be worse off than in the absence of trade liberalization.

Let $\mu(\underline{\varepsilon}^{\lambda-}, \tau)$ be the share of sectors having been in state $\underline{\varepsilon}$ for a period shorter than λ , $\mu(\underline{\varepsilon}^{\lambda+}, \tau)$ the share of sectors having been in state $\underline{\varepsilon}$ for a period longer than λ , and $\mu(\bar{\varepsilon}, \tau)$ the share of sectors in state $\bar{\varepsilon}$. These shares are then given by

$$\begin{aligned}\mu(\underline{\varepsilon}^{\lambda-}, \tau) &= \begin{cases} \frac{1}{2} & \text{if } \tau \leq \lambda \\ \frac{1}{2} \frac{1-e^{-\lambda}}{1-e^{-\tau}} & \text{if } \tau > \lambda \end{cases} \\ \mu(\underline{\varepsilon}^{\lambda+}, \tau) &= \begin{cases} 0 & \text{if } \tau \leq \lambda \\ \frac{1}{2} \frac{e^{-\lambda}-e^{-\tau}}{1-e^{-\tau}} & \text{if } \tau > \lambda \end{cases} \\ \mu(\bar{\varepsilon}, \tau) &= \frac{1}{2}.\end{aligned}$$

It is easily seen that if $\lambda = 0$, then $\mu(\underline{\varepsilon}^{\lambda-}, \tau) = 0$ and $\mu(\underline{\varepsilon}^{\lambda+}, \tau) = \frac{1}{2}$, and if $\lambda \rightarrow \infty$, then $\mu(\underline{\varepsilon}^{\lambda-}, \tau) = \frac{1}{2}$ and $\mu(\underline{\varepsilon}^{\lambda+}, \tau) = 0$. For any $\lambda \in (0, \infty)$, it is, however, the case that the longer an agreement has existed, the lower will the share of sectors having been in state $\underline{\varepsilon}$ for a period shorter than λ be, and the larger will the share of sectors having been in state $\underline{\varepsilon}$ longer than λ be. Letting $\tau \rightarrow \infty$, $\mu(\underline{\varepsilon}^{\lambda-}, \tau)$ and $\mu(\underline{\varepsilon}^{\lambda+}, \tau)$ will converge to $\frac{1-e^{-\lambda}}{2}$ and $\frac{e^{-\lambda}}{2}$, respectively. These changes in $\mu(\underline{\varepsilon}^{\lambda-}, \tau)$ and $\mu(\underline{\varepsilon}^{\lambda+}, \tau)$ will have two effects. On the one hand, the risk of being exposed to the situation where further protection would be desirable but is no longer possible (i.e. being in state $\underline{\varepsilon}^{\lambda+}$), increases. On the other hand, the likelihood of being in the position of being subject to a trading partner's protection although this is undesirable (i.e. being in state $\bar{\varepsilon}$ and exposed to the trading partner's corresponding sector being in state $\underline{\varepsilon}^{\lambda-}$) decreases.

Define

$$\Pi \equiv \frac{w_C(\bar{\varepsilon}) - w_N(\bar{\varepsilon})}{w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})}.$$

Being the ratio between gains and losses from liberalization in the two states, Π can be regarded as a measure of the expected relative benefit from liberalizing trade. In fact, Π can be seen as an indicator of whether liberalization is worthwhile in the absence of any safeguard provisions. If $\Pi < 1$, the losses from liberalization in the bad state exceed the gains in the good state and hence, no liberalization is preferable to liberalization. And if $\Pi > 1$, liberalization will be beneficial overall, even

if no safeguards are provided, because the gains from liberalization in the good state are larger than the losses from liberalization in the bad state.

The average sector payoff \bar{w} at time τ for a given λ is given by

$$\begin{aligned}\bar{w}(\tau, \lambda) = & \mu(\bar{\varepsilon}, \tau) \{ \mu(\underline{\varepsilon}^{\lambda^-}, \tau) w_N(\bar{\varepsilon}) + [1 - \mu(\underline{\varepsilon}^{\lambda^-}, \tau)] w_C(\bar{\varepsilon}) \} \\ & + \mu(\underline{\varepsilon}^{\lambda^-}, \tau) w_N(\underline{\varepsilon}) \\ & + \mu(\underline{\varepsilon}^{\lambda^+}, \tau) \{ \mu(\underline{\varepsilon}^{\lambda^-}, \tau) w_N(\underline{\varepsilon}) + [1 - \mu(\underline{\varepsilon}^{\lambda^-}, \tau)] w_C(\underline{\varepsilon}) \}\end{aligned}$$

Plugging in the value for μ derived above, $\bar{w}(\tau, \lambda)$ can be expressed as follows

$$\bar{w}(\tau, \lambda) = \begin{cases} \frac{1}{2} w_N(\underline{\varepsilon}) + \frac{1}{4} w_N(\bar{\varepsilon}) + \frac{1}{4} w_C(\bar{\varepsilon}) & \text{if } \tau < \lambda \\ \frac{1}{2} w_C(\underline{\varepsilon}) + \frac{1}{2} w_C(\bar{\varepsilon}) \\ + \frac{1-e^{-\lambda}}{4(1-e^{-\tau})} \left(3 - \frac{1-e^{-\lambda}}{1-e^{-\tau}} - \Pi \right) [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] & \text{if } \tau \geq \lambda \end{cases}. \quad (1)$$

It immediately follows that

$$\begin{aligned}\bar{w}(\tau, 0) &= \frac{1}{2} w_C(\underline{\varepsilon}) + \frac{1}{2} w_C(\bar{\varepsilon}) \\ \lim_{\lambda \rightarrow \infty} \bar{w}(\tau, \lambda) &= \frac{1}{2} w_N(\underline{\varepsilon}) + \frac{1}{4} w_N(\bar{\varepsilon}) + \frac{1}{4} w_C(\bar{\varepsilon}).\end{aligned}$$

Hence, average sector payoffs will be constant over time in the absence of a safeguard or when the safeguard can be indefinitely applied.

For a given λ , the average discounted flow of future payoffs \bar{v} at time τ is obtained as follows

$$\bar{v}(\tau, \lambda) = \int_0^{\infty} \bar{w}(\tau + s, \lambda) e^{-\delta s} ds.$$

Since \bar{w} is constant over time when $\lambda = 0$ or $\lambda = \infty$, the same is true for the average discounted flow of future payoffs, which in these cases becomes

$$\begin{aligned}\bar{v}(\tau, 0) &= \frac{1}{2\delta} w_C(\underline{\varepsilon}) + \frac{1}{2\delta} w_C(\bar{\varepsilon}) \\ \lim_{\lambda \rightarrow \infty} \bar{v}(\tau, \lambda) &= \frac{1}{2\delta} w_N(\underline{\varepsilon}) + \frac{1}{4\delta} w_N(\bar{\varepsilon}) + \frac{1}{4\delta} w_C(\bar{\varepsilon}).\end{aligned}$$

For $\lambda \in (0, \infty)$ it is, however, the case that \bar{w} changes for $\tau > \lambda$. The following expression for $\bar{v}(\tau, \lambda)$ can be derived (see the Appendix).

$$\bar{v}(\tau, \lambda) = \begin{cases} \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] \\ + \frac{(\Pi - 1 - e^{-\lambda})e^{-(1+\delta)\lambda}e^{\delta\tau}}{4\delta[1+\delta(1-e^{-\lambda})]} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] & \text{if } \tau < \lambda \\ \frac{1}{2\delta} [w_C(\underline{\varepsilon}) + w_C(\bar{\varepsilon})] \\ + \frac{1-e^{-\lambda}}{4\delta[1+\delta(1-e^{-\tau})]} [2 + 3\delta + e^{-\lambda} - \delta\frac{1-e^{-\lambda}}{1-e^{-\tau}} \\ - (1 + \delta)\Pi] [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] & \text{if } \tau \geq \lambda \end{cases}. \quad (2)$$

4.2 Optimization

Ex ante, given a degree of trade liberalization, a government will choose the adjustment phase length maximizing the average discounted flow of future sector payoffs at $\tau = 0$, which by (2) is given by

$$\begin{aligned} \bar{v}(0, \lambda) &= \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] \\ &\quad + \frac{(\Pi - 1 - e^{-\lambda})e^{-(1+\delta)\lambda}}{4\delta[1 + \delta(1 - e^{-\lambda})]} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})]. \end{aligned}$$

Let $\hat{\lambda} \equiv \arg \max_{\lambda \in [0, \infty]} \bar{v}(0, \lambda)$ be the ex ante optimal adjustment phase.

The following proposition relates the optimal choice of adjustment phase length to the relative benefit from introducing trade liberalization.

Proposition 1 *The optimal adjustment phase length unambiguously decreases in the ratio between the gains and losses from liberalization. More specifically*

$$\hat{\lambda} = \begin{cases} \infty & \text{if } \Pi \leq 1 \\ \ln \left[\frac{2\delta(1+\delta)}{2+3\delta+\Pi\delta^2 - \sqrt{(2+3\delta+\Pi\delta^2)^2 - 4(\Pi-1)\delta(1+\delta)^3}} \right] & \text{if } \Pi \in (1, 2 + \frac{1}{1+2\delta}) \\ 0 & \text{if } \Pi \geq 2 + \frac{1}{1+2\delta} \end{cases}. \quad (3)$$

Proof. See the Appendix. ■

The intuition behind this result is straightforward. The larger is the ratio of gains to losses from liberalization, the less should liberalization be inhibited. Hence, while not constraining the use of the safeguard

is optimal if the gains are outweighed by the losses from liberalization ($\Pi \leq 1$)¹⁴, not having any safeguard provision is optimal if the gains are sufficiently larger than the losses from liberalization ($\Pi \geq 2 + \frac{1}{1+2\delta}$). For intermediate values of the ratio of gains to losses from liberalization ($\Pi \in (1, 2 + \frac{1}{1+2\delta})$), a safeguard with a time limit is optimal.¹⁵

It is also important to stress that this result does not hinge on political economic or moral hazard considerations, which might potentially add further incentives for limiting the length of the adjustment phase. An industry in decay (in the bad state) enjoying the protection of a safeguard should have low incentives to take measures to become more competitive (reach the good state). In fact, it might rather lobby for continued protection than take any such steps. Hence, a government might prefer to have an outside commitment by means of a trade agreement, which limits its possibilities of delivering protection. Letting switches between states be exogenously determined, the present model does not take such considerations into account, however.

It is important to emphasize that the derived results hinge on the assumption that ex ante, in negotiations about a trade agreement, the two countries decide to adopt a clear-cut rule governing for how long a safeguard can be applied, and then adhere to it. Thus, it is implicitly assumed that mutually agreed-upon breaches against this rule will not take place, once the agreement has been implemented. Hence, the derived solutions are optimal, subject to the constraint that such a rule is applied. In fact, the derived solutions may not be first-best due to this constraint (see next subsection).

The following lemma demonstrates that the optimal length of the adjustment phase decreases in the discounting rate.

Lemma 1 *If $\hat{\lambda} \in (0, \infty)$, then $\hat{\lambda}$ decreases in δ . In particular, $\hat{\lambda}$ may become zero as δ is increased.*

Proof. See the Appendix. ■

Thus, the more future payoffs are discounted, the shorter the optimal length of the adjustment phase will become, and the lower the threshold value for the ratio between gains and losses from liberalization, above

¹⁴As previously noted, no liberalization would actually be preferred to liberalization in the absence of the safeguard if $\Pi \leq 1$.

¹⁵A more thorough interpretation of this result can be found in the next subsection.

which no safeguard is the optimal solution, will be. Intuitively, this result is not straightforward. Actually, it is not clear why the discount rate should have any impact at all. What drives this result, however, is the fact that a higher discount rate implies that the prospect of having a sector in state $\underline{\varepsilon}^{\lambda+}$ sometime in the future carries less weight. Hence, the negative impact of a shorter adjustment phase on this category of sectors is given less weight, thereby moving the optimal adjustment phase length solution in favor of what is optimal for sectors in state $\bar{\varepsilon}$.

From an analytical point of view, the case when future payoffs carry the same weight as instant payoffs is of interest.¹⁶ Letting δ approach zero, $\hat{\lambda} = \ln(\frac{2}{\Pi-1})$ for $\Pi \in (1, 3)$. Hence, the qualitative result in proposition 1, with finite solutions for $\hat{\lambda}$ for intermediate values of Π , is valid even as the weight given to future payoffs approximates the weight given to instant payoffs.

4.3 Ex ante Suboptimality

As previously emphasized, having a rule that is supposed to apply in all contingencies may act as a constraint. Optimality under such a rule may not yield the first-best solution, which is confirmed by the following proposition.

Proposition 2 *The optimal solution given by (3) is first-best if and only if $\Pi \leq 1$.*

Proof. Consider any point in time after the agreement has been implemented. There will exist four categories of sectors, depending on the own state and the state of the corresponding sector in the other country. The first category comprises corresponding sectors in state $\bar{\varepsilon}$ in both countries, in which case no safeguard is applied. Thus, this category need not be considered. If the corresponding sectors in the two countries are both in state $\underline{\varepsilon}$, both are in need of the safeguard, irrespective of how long they have been in that state. Hence, no time limit ($\lambda = \infty$) on the use of the safeguard is optimal for this category. The remaining two categories comprise corresponding sectors in different states in the two countries. It is straightforward that applying the safeguard is globally

¹⁶Since the value of discounted future flows of payoffs would be infinitely large for any λ in this case, it is not possible to analytically solve the optimization problem. Letting the discount rate go to zero nevertheless renders insights into how the optimal adjustment phase length is affected, if future payoffs are given approximately the same weight as present payoffs.

optimal if $\Pi < 1$, while not applying it is globally optimal if $\Pi > 1$ (and aggregate sector payoffs are the same with or without a safeguard, if $\Pi = 1$). Thus, for these two categories, no time limit ($\lambda = \infty$) on the use of the safeguard is optimal if $\Pi < 1$, while not allowing the safeguard to be applied at all ($\lambda = 0$) is optimal if $\Pi > 1$ (and any $\lambda \in [0, \infty]$ is optimal if $\Pi = 1$). It immediately follows that under a rule applied under all contingencies, the optimal solution given by (3) is first-best if and only if $\Pi \leq 1$. ■

When $\Pi \leq 1$, the above derived optimal solution is thus first-best. Moreover, there will exist no ex post incentives to deviate from the agreement, because losses from liberalization will never be incurred, while gains from liberalization will be made in one-quarter of trades (i.e. between any pair of country sectors that are both in the good state).

The categorization of sectors in the proof of the above proposition helps us interpret the somewhat surprising result in proposition 1 that the optimal adjustment phase length is gradually decreasing in Π , rather than jumping from infinity to zero at $\Pi = 1$. The underlying reason is that, under a clear-cut rule, there will be conflicting interests between the different categories of sectors when $\Pi > 1$. In the categories comprising corresponding sectors in different states in the two countries, it is globally optimal not to apply the safeguard, while in the category where corresponding sectors are simultaneously in the bad state, it is optimal to implement the safeguard. The optimal adjustment phase length obtained in proposition 1 thus strikes a balance between gains and losses made by these different categories of sectors under a safeguard with a predetermined time limit. As Π increases above one, the gain from not applying the safeguard in the two categories of corresponding sectors in different states increases. The optimal solution will thus increasingly be tilted in favor of the globally optimal solution for these two categories, i.e. $\hat{\lambda}$ will fall, as Π rises. Eventually, when $\Pi \geq 2 + \frac{1}{1+2\delta}$, the gain for these two categories from having no safeguard outweighs the loss incurred for corresponding sectors simultaneously in the bad state and not allowed to deviate and hence, $\hat{\lambda} = 0$ is optimal.

It immediately follows from the preceding discussion why the optimal solution is not first-best whenever $\Pi > 1$. The optimal solution can be seen as a compromise between the various categories of sectors. When $\Pi \in (1, 2 + \frac{1}{1+2\delta})$, it is the case that $\hat{\lambda} \in (0, \infty)$ and thus, the safeguard will be applied if and only if at least one of any pair of corresponding sectors finds itself in state $\underline{\epsilon}^{\lambda^-}$. This is globally suboptimal, both for the two categories of corresponding sectors in different states, for which no

safeguard at all would be optimal, and for corresponding sectors that are simultaneously in state $\underline{\varepsilon}^{\lambda+}$, in which case both countries would be better off scaling back liberalization, but neither of them is permitted to do so. When $\Pi \geq 2 + \frac{1}{1+2\delta}$ and hence $\hat{\lambda} = 0$, global suboptimality only arises in the latter category, i.e. whenever corresponding sectors are simultaneously in state $\underline{\varepsilon}$.

The proof of the previous proposition suggests the first-best solution to be of bang-bang type. However, this is only true with some qualifications for $\Pi > 1$. The first-best solution when $\Pi > 1$ prescribes the safeguard to be used whenever corresponding sectors are simultaneously in the bad state, while not allowing for it to be applied when they are in different states. Hence, the safeguard should be applied if and only if and for as long as both countries would agree on this.

There are, however, strong practical and political reasons for assuming away the possibility of allowing for a safeguard to be applied, if and only if it has unanimous support. First, it will require the immediate withdrawal of the safeguard as soon as required by one country. More specifically, the implementability of the first-best solution will rest on liberalization being suspended only as long as both sectors are in the bad state; as soon as one country's sector switches to the good state, liberalization must be reintroduced, although that will make the country whose sector is still in the bad state worse off. The duration of protection under a safeguard being dependent on the state abroad will make protection unpredictable. In the present context of entirely exogenously determined switches in states, this may be of minor importance. In reality, it could be politically difficult to agree upon such a rule. One of the benefits of having a clear-cut rule governing the length of the adjustment phase is that it makes the future predictable. A sector in need of protection will be able to relate its adjustment measures to the time limits of the protection granted.

Second, it might be difficult to implement the first-best solution in a multi-country framework. The first-best solution might prescribe the safeguard to be applied only against countries with corresponding sectors also in the bad state, while countries with corresponding sectors in the good state would not be affected. This discriminatory use of the safeguard would amount to a breach against one of the fundamental principles of all trade agreements, the MFN clause.¹⁷ In contrast, the

¹⁷For rigorous theoretical support of the MFN principle, I refer to Bagwell and Staiger (1999) and Bagwell and Staiger (2002). Horn and Mavroidis (2001) provide a survey of economic and legal aspects of the MFN clause.

derived optimal solution has the advantage of being easily reconciled with the MFN principle. Besides, the first-best solution might require coordination and supervision in a multi-country framework. A supranational monitoring or enforcement agency might be necessary, which might be costly and probably politically unfeasible. Apart from strong practicability constraints, it is highly unlikely that sovereign states would ever cede their control over trade policy to a supranational agency. The fact that countries have agreed to adopt rules governing the conduct of various provisions rather than letting trade policy be supranationally fine-tuned renders support to the underlying presumption that first-best solutions may not be feasible.

5 Ex post Implications of a Time Limit

5.1 Instant Payoffs under Optimality

When the optimal adjustment phase length is determined, the flow of future payoffs from the moment the agreement is implemented is maximized. As shown in the previous section, the optimal solution is having a safeguard without any upper time limit on its use if $\Pi \leq 1$, and including no safeguard provision at all if $\Pi \geq 2 + \frac{1}{1+2\delta}$. In these two cases, the average instant payoff \bar{w} is constant over time. If, however, $\Pi \in (1, 2 + \frac{1}{1+2\delta})$, the optimal solution prescribes a safeguard with an upper time limit on its use, i.e. $\hat{\lambda} \in (0, \infty)$. In this case, both the average instant payoff \bar{w} and the average discounted value of the flow of average future payoffs \bar{v} will change over time, once the agreement has been implemented. First, the effect on instant payoffs will be assessed.

Lemma 2 *If $\Pi \in (1, 2 + \frac{1}{1+2\delta})$, implementing the optimal adjustment phase length $\hat{\lambda}$ results in \bar{w} being constant for $\tau \leq \hat{\lambda}$. As τ increases beyond $\hat{\lambda}$, \bar{w} initially increases, but eventually decreases, thereby converging to a value unambiguously larger than its value at the agreement's inception ($\lim_{\tau \rightarrow \infty} \bar{w}(\tau, \hat{\lambda}) > \bar{w}(0, \hat{\lambda})$).*

Proof. See the Appendix. ■

To get an intuitive understanding of this result, it is necessary to determine in which ways the average payoff $\bar{w}(\tau, \lambda)$ is affected over time. Let $\mu \equiv \mu(\underline{\varepsilon}^{\lambda-}, \tau)$, the share of sectors having been in state $\underline{\varepsilon}$ for a period shorter than λ . From the previous section, we know that μ equals one-half for $\tau \leq \hat{\lambda}$, decreases monotonically in τ for $\tau > \hat{\lambda}$, and converges to $\frac{1-e^{-\hat{\lambda}}}{2}$ as τ goes to infinity. There are three ways in which this change

in μ over time has an impact on the average sector payoff. First, it unambiguously increases average payoffs in state $\bar{\varepsilon}$ for $\tau > \hat{\lambda}$. Second, it unambiguously reduces average payoffs in state $\underline{\varepsilon}^{\hat{\lambda}+}$ for $\tau > \hat{\lambda}$. Third, it unambiguously depresses the total average payoff as the share of sectors in state $\underline{\varepsilon}^{\hat{\lambda}+}$ increases and the share of sectors in state $\underline{\varepsilon}^{\hat{\lambda}-}$ decreases correspondingly. This negative effect due to the shift of sectors in state $\underline{\varepsilon}^{\hat{\lambda}-}$ to state $\underline{\varepsilon}^{\hat{\lambda}+}$ is exacerbated by the second effect that the average payoff in state $\underline{\varepsilon}^{\hat{\lambda}+}$ decreases. To understand the impact of these different effects on \bar{w} over time, \bar{w} can be expressed as follows

$$\begin{aligned}\bar{w}(\tau, \lambda) &= \frac{1}{2}[\mu w_N(\bar{\varepsilon}) + (1 - \mu)w_C(\bar{\varepsilon})] \\ &\quad + \mu w_N(\underline{\varepsilon}) + \left(\frac{1}{2} - \mu\right)[\mu w_N(\underline{\varepsilon}) + (1 - \mu)w_C(\underline{\varepsilon})] \\ &= \frac{1}{2}\{w_C(\bar{\varepsilon}) - \mu\Pi[w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})]\} \\ &\quad + \frac{1}{2}w_N(\underline{\varepsilon}) - (1 - \mu)\left(\frac{1}{2} - \mu\right)[w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})].\end{aligned}$$

The first effect enters via the first term, while the second and the third effects enter through the third term. While the first, positive effect is linear in μ , the combined impact of the last two, negative effects is quadratic in μ . It is easily established that the net marginal effect of a decrease in μ below one-half (i.e. at time $\tau = \hat{\lambda}$) is strictly positive. Hence, \bar{w} will increase beyond $\hat{\lambda}$. As time passes and μ decreases further, the marginal impact on the third term becomes increasingly negative, since this term is quadratic in μ . Eventually, it will exceed the positive, constant marginal impact on the first term, and the net marginal effect will therefore become negative and \bar{w} will start falling. Thus, the combined effect of bad-state sectors in need of extended protection becoming increasingly worse off on the one hand, and a shift of sectors from those protected by a safeguard to those that are not on the other hand, will eventually lead to a decline in the average sector payoff.

5.2 Flows of Future Payoffs under Optimality

The lemma of the previous subsection, showing that under the optimal adjustment phase length $\hat{\lambda}$ instant payoffs are constant for $\tau \leq \hat{\lambda}$, initially increase and then decrease as τ increases beyond $\hat{\lambda}$, suggests that the discounted value of future flows of payoff should initially increase and eventually fall. The following lemma demonstrates that the eventual decrease in \bar{v} takes place after the agreement has been in place for a longer time than the adjustment phase length.

Lemma 3 *If $\Pi \in (1, 2 + \frac{1}{1+2\delta})$, implementing the optimal adjustment phase length $\widehat{\lambda}$ results in \bar{v} increasing beyond $\widehat{\lambda}$, but eventually decreasing, thereby converging to a value unambiguously larger than its value at the agreement's inception ($\lim_{\tau \rightarrow \infty} \bar{v}(\tau, \widehat{\lambda}) > \bar{v}(0, \widehat{\lambda})$).*

Proof. See the Appendix. ■

The next lemma follows immediately.

Lemma 4 *For an agreement with an optimal adjustment phase length $\widehat{\lambda}$, the following is true:*

- (i) $\bar{v}(\tau, \widehat{\lambda}) > \bar{v}(\tau, 0)$ for all τ
- (ii) $\bar{v}(\tau, \widehat{\lambda}) > \bar{v}(\tau, \infty)$ for all τ
- (iii) $\bar{v}(\tau, \widehat{\lambda}) > \bar{v}_N(\tau)$ for all τ .

Proof. (i) and (ii) Since $\bar{v}(0, \widehat{\lambda}) > \bar{v}(0, \lambda)$ for all $\lambda \neq \widehat{\lambda}$ ($\widehat{\lambda}$ maximizes $\bar{v}(0, \lambda)$), $\bar{v}(\tau, \widehat{\lambda}) > \bar{v}(0, \widehat{\lambda})$ for $\tau > 0$ (see lemma 3) and $\bar{v}(\tau, \lambda)$ is constant over time for $\lambda = 0$ or $\lambda = \infty$, it immediately follows that $\bar{v}(\tau, \widehat{\lambda}) > \bar{v}(\tau, 0)$ and $\bar{v}(\tau, \widehat{\lambda}) > \bar{v}(\tau, \infty)$ for all τ .

(iii) Since $\bar{v}(\tau, \infty) = \frac{1}{4\delta}[w_C(\bar{\varepsilon}) + w_N(\bar{\varepsilon}) + 2w_N(\underline{\varepsilon})] > \frac{1}{2\delta}[w_N(\bar{\varepsilon}) + w_N(\underline{\varepsilon})] = \bar{v}_N(\tau)$, it immediately follows from (ii) that $\bar{v}(\tau, \widehat{\lambda}) > \bar{v}(\tau, 0)$ for all τ . ■

The first claim of the above proposition implies that a safeguard with time limit $\widehat{\lambda}$ is preferred to having no safeguard at all over the entire time horizon. Thus the optimal time limit $\widehat{\lambda}$ at any point in time yields a higher discounted value of future payoffs than having no safeguard at all, which is equivalent to letting countries deal with losses from liberalization domestically through redistribution. The optimal time limit on the use of the safeguard is however such that the losses incurred through the use of the safeguard are always outweighed by its gains vis-à-vis the case of having no safeguard at all over the entire time horizon.

The second claim implies that a safeguard with time limit $\widehat{\lambda}$ is preferred to having a safeguard without a time limit over the entire time horizon. Hence the optimal time limit $\widehat{\lambda}$ at any point in time yields a higher discounted value of future payoffs than having a safeguard without any time limit, which is equivalent to letting countries deal with all losses from liberalization externally through the scaling back of liberalization. The underlying reason is that there are gains to be made from limiting

the use of the safeguard whenever it is globally harmful, i.e. whenever one of the corresponding sectors finds itself in state $\bar{\varepsilon}$. However, there will also be losses incurred whenever it would globally be more efficient to extend the use of the safeguard, i.e. whenever the corresponding sectors in both countries find themselves in state $\underline{\varepsilon}^{\lambda+}$. But the optimal time limit on the use of the safeguard ensures that, at any point in time, these losses are always outweighed by the gains of restricting the use of the safeguard, as compared to the case of having a safeguard with no time limit at all.

The third claim, finally, implies that a safeguard with time limit $\hat{\lambda}$ is preferred to no agreement at all over the entire time horizon.

To summarize, the discounted value of future flows of payoffs will always be larger than in the absence of any agreement, or if no safeguard or a safeguard with no time limit is applied. However, the fact that \bar{v} will eventually fall suggests that the agreed-upon time limit on the use of the safeguard might no longer be optimal ex post, thus leaving room for ex post adjustments. In what follows, it will be demonstrated that this is indeed the case.

5.3 Ex post Suboptimality

Since the discounted value of future flows of payoffs \bar{v} is constant over time if no safeguard is included in the agreement ($\hat{\lambda} = 0$), it immediately follows that if $\hat{\lambda} = 0$ is optimal at the agreement's inception, it will be so over the infinite time horizon. The same applies to the case when a safeguard with no time limit ($\hat{\lambda} = \infty$) is implemented. For $\hat{\lambda} \in (0, \infty)$, however, \bar{v} will vary over time and it may be the case that $\hat{\lambda}$ is no longer optimal, once the agreement has been implemented.

By introspection of (2), it is easily seen that the optimal solution is independent of τ for $\tau \leq \lambda$. Hence, the optimal solution is also ex post optimal as long as $\tau \leq \lambda$. For $\tau > \lambda$, this is no longer the case, however, as demonstrated by the following proposition.

Proposition 3 *If $\Pi \in (1, 2 + \frac{1}{1+2\delta})$, the ex post optimal solution for λ will eventually be higher than the agreed-upon solution $\hat{\lambda}$. As τ increases, so will the ex post optimal solution for λ .*

Proof. See the Appendix. ■

Eventually, countries would be better off, if a higher λ had been chosen at the agreement's inception.¹⁸ The intuition behind this result is that when the agreement is implemented, insufficient weight is attributed to the category of sectors that have been in the bad state for longer than the length of the adjustment phase and hence, are in need of further protection. Since the share of this category of sectors is zero in the initial phase of the agreement, i.e. as long as $\tau \leq \hat{\lambda}$, and then increases, the weight it is given is too small ex ante.

It is important to emphasize that the ex post optimal adjustment phase length is optimal in the sense that it would yield the highest value of discounted flows of future payoffs at a specific point in time. A government will thus ex post perceive that a higher value of discounted future payoff flows could have been attained if a different adjustment phase length had been chosen. Hence, the ex ante suboptimality associated with including a clear-cut rule prescribing an upper time limit on the use of the safeguard will also lead to ex post suboptimality.

The agreed-upon time limit for the use of the safeguard will thus eventually be perceived to be too short. Or, in other words, dissatisfaction will grow over time. The demands for increasing the time limit might increase, which implies that there may be some room for ex post renegotiation to modify the adjustment phase length.¹⁹

It is, however, important to emphasize that the optimal solution derived in the previous section implicitly rests on the assumption that ex post readjustments are not possible, or in other words, that governments can commit to the rules of the agreement. If the time limit could be modified ex post, governments should correctly anticipate such adjustments ex ante, which would feed back into the determination of the time limit being implemented at the inception of the agreement. In fact, if readjustments were possible, an optimal solution might prescribe adjustment taking place such that optimality would be satisfied at any point in time. Such solutions are obviously an interesting topic of future research.

¹⁸In the proof of proposition 4, it is shown that the ex post optimal time limit will actually be strictly smaller than $\hat{\lambda}$ as τ increases beyond $\hat{\lambda}$. While the observation that countries will initially perceive $\hat{\lambda}$ to be too large as τ increases beyond $\hat{\lambda}$ is interesting, it will not be addressed any further. From a practical point of view, there are likely to be obstacles to ex post lower the adjustment phase length. Anyway, the ex post optimal time limit will eventually be increasingly larger than $\hat{\lambda}$.

¹⁹In the present model, where countries are equally well off due to a continuum of sectors, getting unanimous support for any ex post readjustments would be no problem. If the number of sectors were finite, however, countries might not be equally well off and hence, agreement on ex post readjustments might be harder to achieve.

More generally, however, the fact that ex post suboptimality will arise in the presence of a finite, time limit on the use of safeguards implies that the obtained solution is not time consistent. Ex post, i.e. for $\tau > \widehat{\lambda}$, the originally adopted solution will no longer yield the highest possible value of flows of discounted future payoffs and hence, a change in the time limit would be beneficial.²⁰

The ex post suboptimality identified here rests on the assumption of a continuum of sectors. If the number of sectors were instead finite, aggregate payoffs would differ from the ex ante expected value and hence, ex post dissatisfaction would not necessarily arise. However, since the derived expressions for \bar{v} would still apply as expected values, it could ex ante be expected that ex post suboptimality might arise.

The result of ex post suboptimality of a time limit on being exempted from a cooperative arrangement renders an interesting interpretation of the ongoing controversy about the growth and stability pact in the euro area, an agreement under which participating countries are obliged to adhere to certain prespecified rules. In particular, the pact prescribes budget deficits not to exceed 3% of GDP in more than three consecutive years. At the time when the pact was negotiated, the prospect of a country finding itself in the situation of needing to run a deficit exceeding 3% of GDP in a fourth consecutive year was in the distant future and therefore, low weight was attributed to this possibility. It can be argued that while the budget deficit rule seemed optimal when the pact was signed, it is increasingly being regarded as too rigid. In fact, during recent years, several participating countries have found themselves in the position of risking to run a deficit exceeding 3% of GDP for more than three years in a row. Not surprisingly, voices have been raised for relaxing this rule, or at least interpreting it more generously.

6 Conclusion

This paper has attempted to shed light on the determination of time limits for how long countries should be allowed to withdraw commitments made under a trade agreement. A symmetric two-country model, where sectors stochastically switch between two states, has been applied. While there are gains to be made from liberalization in the good state, losses will be incurred under liberalization in the bad state. It may therefore

²⁰The present time inconsistency seems to originate in having a rigid rule prescribing a fixed time limit to being exempted from a commitment, thereby suggesting a similarity to the time inconsistency identified in Kydland and Prescott (1977).

be desirable to agree on a rule, prescribing how long protection may be granted to sectors in bad states. It has been shown that the optimal time limit on protection will depend on the ratio between the gain from liberalization under the good state and the loss from liberalization under the bad state. For low ratios, no upper limit on protection is optimal and for high ratios, allowing for no protection at all is optimal. For intermediate ratios, however, a strictly positive and finite time limit is optimal. The optimal time limit thus serves to balance the interests of winners and losers in liberalization across countries.

Whenever the ex ante agreed-upon time limit is strictly positive and finite, it will eventually be perceived as being too low to an increasing extent. Thus, countries will ex post find the agreed upon adjustment phase length suboptimal. Hence, the dissatisfaction with the agreement will grow over time.

The present analysis has entirely focused on the optimal length of a safeguard allowing for the withdrawal of liberalization commitments. No link to the actual extent of these commitments was established. An important topic for further research is to relate the optimal adjustment phase length to the degree of liberalization agreed-upon in trade negotiations. For that purpose, it is necessary to apply a more specific model, from which the interrelation between liberalization and the associated potential gains and losses can be derived. If, for instance, the ratio between potential gains and losses from liberalization will decrease in the degree of liberalization, the optimal adjustment phase length will increase in the degree of liberalization. Such an outcome would correspond well to the common perception that, by increasing the exposure to world markets, more liberalization ought to be combined with more flexibility, in particular increasing possibilities to scale back liberalization, for example through more generous time limits on safeguards.

7 References

Bagwell, Kyle and Staiger, Robert W. (1990), *A Theory of Managed Trade*, American Economic Review, 80(4), 779-795

Bagwell, Kyle and Staiger, Robert W. (1999), *An Economic Theory of GATT*, American Economic Review, 89(1), 215-248

Bagwell, Kyle and Staiger, Robert W. (2002), *The Economics of the World Trading System*, The MIT Press, Cambridge, Massachusetts

Ethier, Wilfried J. (2001), *Theoretical Problems in Negotiating Trade Liberalization*, European Journal of Political Economy, 17, 209-232

Ethier, Wilfried J. (2002), *Escape and Entry Mechanisms in the Multilateral Trade System*, PIER Working Paper 02-009, Department of Economics, University of Pennsylvania

Feenstra, Robert C. (1987), *Incentive Compatible Trade Policies*, Scandinavian Journal of Economics, 89(3), 373-387

Feenstra, Robert C. and Lewis, Tracy R. (1991), *Negotiated Trade Restrictions With Private Political Pressure*, Quarterly Journal of Economics, 106(4), 1287-1307

Herzing, Mathias (2005a), *Can Self-Destructive Trade Agreements Be Optimal?*, mimeo, Institute for International Economic Studies, Stockholm University, Sweden

Herzing, Mathias (2005b), *Escape and Optimal Compensation under Trade Agreements*, mimeo, Institute for International Economic Studies, Stockholm University, Sweden

Hochman, Gal (2004), *The Interdependence of Safeguard Measures and Tariffs*, mimeo, William Davidson Faculty of Industrial Engineering and Management, Technion, Israel

Horn, Henrik and Mavroidis, Petros C. (2001), *Economic and Legal Aspects of the Most-Favored-Nation Clause*, European Journal of Political Economy, 17, 233-279

Jensen, Richard and Thursby, Marie (1990), *Tariffs with Private Information and Reputation*, Journal of International Economics, 29(1-2), 43-67

Kydland, Finn E. and Prescott, Edward C. (1977), *Rules Rather than Discretion: The Inconsistency of Optimal Plans*, Journal of Political Economy, 85(3), 473-491

Martin, Alberto and Vergote, Wouter (2004), *A Case for Strategic Antidumping*, mimeo, Columbia University, New York

Riezman, Raymond (1991), *Dynamic Tariffs with Asymmetric Information*, Journal of International Economics, 30(3-4), 267-284

Rosendorff, B. Peter and Milner, Helen (2001), *The Optimal Design of International Trade Institutions: Uncertainty and Escape*, International Organization, 55(4), 829-857

World Trade Organization (1994a), *The General Agreement on Tariffs and Trade*, Geneva, Switzerland

World Trade Organization (1994b), *Agreement on Safeguards*, Geneva, Switzerland

8 Appendix

8.1 Derivation of $\bar{v}(\tau, \lambda)$

The expected discounted flow of future payoffs for $\tau < \lambda$ is given by

$$\begin{aligned}
\bar{v}(\tau, \lambda) &= \int_0^\infty \bar{w}(\tau + t)e^{-\delta t} dt \\
&= \int_0^{\lambda-\tau} \frac{1}{4} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] e^{-\delta t} dt \\
&\quad + \int_{\lambda-\tau}^\infty \left\{ \frac{1}{2} w_C(\underline{\varepsilon}) + \frac{1}{2} w_C(\bar{\varepsilon}) \right. \\
&\quad \left. + \frac{1}{4} \frac{1-e^{-\lambda}}{1-e^{-\tau-t}} \left[3 - \frac{1-e^{-\lambda}}{1-e^{-\tau-t}} - \Pi \right] [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \right\} e^{-\delta t} dt \\
&= \frac{1 - e^{-\delta(\lambda-\tau)}}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] + \frac{e^{-\delta(\lambda-\tau)}}{2\delta} [w_C(\underline{\varepsilon}) + w_C(\bar{\varepsilon})] \\
&\quad + \frac{1 - e^{-\lambda}}{4} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \int_{\lambda-\tau}^\infty \left[\frac{3 - \Pi}{1 - e^{-\tau-t}} - \frac{1 - e^{-\lambda}}{(1 - e^{-\tau-t})^2} \right] e^{-\delta t} dt.
\end{aligned}$$

The integral can be transformed as follows.

$$\begin{aligned}
&\int_{\lambda-\tau}^\infty \left[\frac{3 - \Pi}{1 - e^{-\tau-t}} - \frac{1 - e^{-\lambda}}{(1 - e^{-\tau-t})^2} \right] e^{-\delta t} dt \\
&= [3 - \Pi - (1 - e^{-\lambda})] \int_{\lambda-\tau}^\infty \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt - (1 - e^{-\lambda}) \int_{\lambda-\tau}^\infty \frac{e^{-\tau-t} e^{-\delta t}}{(1 - e^{-\tau-t})^2} dt \\
&= [3 - \Pi - (1 - e^{-\lambda})] \int_{\lambda-\tau}^\infty \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt \\
&\quad - (1 - e^{-\lambda}) \left\{ \left[\frac{-e^{-\delta t}}{1 - e^{-\tau-t}} \right]_{\lambda-\tau}^\infty - \delta \int_{\lambda-\tau}^\infty \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt \right\} \quad (*) \\
&= [3 - \Pi - (1 - \delta)(1 - e^{-\lambda})] \int_{\lambda-\tau}^\infty \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt - e^{-\delta(\lambda-\tau)}
\end{aligned}$$

The remaining integral is calculated as follows.

$$\begin{aligned}
\int_{\lambda-\tau}^\infty \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt &= \int_{\lambda-\tau}^\infty e^{-\delta t} \sum_{j=0}^\infty e^{-j(\tau+t)} dt = \sum_{j=0}^\infty \int_{\lambda-\tau}^\infty e^{-\delta t - j(\tau+t)} dt \\
&= \sum_{j=0}^\infty \left[-\frac{e^{-\delta t - j(\tau+t)}}{\delta + j} \right]_{\lambda-\tau}^\infty = \frac{e^{-\delta(\lambda-\tau)}}{\delta} \sum_{j=0}^\infty \frac{\delta e^{-j\lambda}}{\delta + j} \quad (\#) \\
&= \frac{e^{-\delta(\lambda-\tau)}}{\delta} \frac{1}{1 - \frac{\delta e^{-\lambda}}{\delta + 1}} = \frac{1 + \delta}{\delta} \frac{e^{-\delta(\lambda-\tau)}}{1 + \delta(1 - e^{-\lambda})}
\end{aligned}$$

Hence,

$$\begin{aligned}
& \int_{\lambda-\tau}^{\infty} \left[\frac{3-\Pi}{1-e^{-\tau-t}} - \frac{1-e^{-\lambda}}{(1-e^{-\tau-t})^2} \right] e^{-\delta t} dt \\
&= \frac{e^{-\delta(\lambda-\tau)}}{\delta[1+\delta(1-e^{-\lambda})]} \left\{ (1+\delta)[3-\Pi - (1-\delta)(1-e^{-\lambda})] - \delta[1+\delta(1-e^{-\lambda})] \right\} \\
&= \frac{e^{-\delta(\lambda-\tau)}}{\delta[1+\delta(1-e^{-\lambda})]} [(1+\delta)(2-\Pi) + e^{-\lambda}].
\end{aligned}$$

Thus the expected discounted flow of future payoffs for $\tau < \lambda$ is given by

$$\begin{aligned}
\bar{v}(\tau, \lambda) &= \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] + \frac{e^{-\delta(\lambda-\tau)}}{4\delta} (\Pi - 2) [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\
&\quad + \frac{e^{-\delta(\lambda-\tau)}}{4\delta[1+\delta(1-e^{-\lambda})]} (1-e^{-\lambda}) [(1+\delta)(2-\Pi) + e^{-\lambda}] [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\
&= \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] \\
&\quad + \frac{e^{-\delta(\lambda-\tau)}}{4\delta[1+\delta(1-e^{-\lambda})]} \left\{ \begin{array}{l} (\Pi - 2)[1+\delta(1-e^{-\tau})] \\ + (1+\delta)(1-e^{-\lambda})(2-\Pi) \\ + e^{-\lambda}(1-e^{-\lambda}) \end{array} \right\} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\
&= \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] \\
&\quad + \frac{[\Pi - 1 - e^{-\lambda}]e^{-\delta(\lambda-\tau)}e^{-\lambda}}{4\delta[1+\delta(1-e^{-\lambda})]} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})].
\end{aligned}$$

The expected discounted flow of future payoffs for $\tau \geq \lambda$ is given by

$$\begin{aligned}
\bar{v}(\tau, \lambda) &= \int_0^{\infty} \bar{w}(\tau + t) e^{-\delta t} dt \\
&= \int_0^{\infty} \left\{ \begin{array}{l} \frac{1}{2}w_C(\underline{\varepsilon}) + \frac{1}{2}w_C(\bar{\varepsilon}) \\ + \frac{1}{4} \frac{1-e^{-\lambda}}{1-e^{-\tau-t}} \left[3 - \frac{1-e^{-\lambda}}{1-e^{-\tau-t}} - \Pi \right] [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \end{array} \right\} e^{-\delta t} dt \\
&= \frac{1}{2\delta} [w_C(\underline{\varepsilon}) + w_C(\bar{\varepsilon})] \\
&\quad + \frac{1-e^{-\lambda}}{4} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \int_0^{\infty} \left[\frac{3-\Pi}{1-e^{-\tau-t}} - \frac{1-e^{-\lambda}}{(1-e^{-\tau-t})^2} \right] e^{-\delta t} dt.
\end{aligned}$$

The integral can be transformed as follows.

$$\int_0^{\infty} \left[\frac{3-\Pi}{1-e^{-\tau-t}} - \frac{1-e^{-\lambda}}{(1-e^{-\tau-t})^2} \right] e^{-\delta t} dt$$

$$\begin{aligned}
&\stackrel{(*)}{=} [3 - \Pi - (1 - e^{-\lambda})] \int_{\lambda-\tau}^{\infty} \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt \\
&\quad - (1 - e^{-\lambda}) \left\{ \left[\frac{-e^{-\delta t}}{1 - e^{-\tau-t}} \right]_0^{\infty} - \delta \int_{\lambda-\tau}^{\infty} \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt \right\} \\
&= [3 - \Pi - (1 - \delta)(1 - e^{-\lambda})] \int_{\lambda-\tau}^{\infty} \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt - \frac{1 - e^{-\lambda}}{1 - e^{-\tau}}
\end{aligned}$$

The remaining integral is calculated as follows.

$$\begin{aligned}
\int_0^{\infty} \frac{e^{-\delta t}}{1 - e^{-\tau-t}} dt &\stackrel{(\#)}{=} \sum_{j=0}^{\infty} \left[-\frac{e^{-\delta t - j(\tau+t)}}{\delta + j} \right]_0^{\infty} = \frac{1}{\delta} \sum_{j=0}^{\infty} \frac{\delta e^{-j\tau}}{\delta + j} \\
&= \frac{1}{\delta} \frac{1}{1 - \frac{\delta e^{-\tau}}{\delta + 1}} = \frac{1 + \delta}{\delta} \frac{1}{1 + \delta(1 - e^{-\tau})}
\end{aligned}$$

Hence,

$$\begin{aligned}
&\int_0^{\infty} \left[\frac{3 - \Pi}{1 - e^{-\tau-t}} - \frac{1 - e^{-\lambda}}{(1 - e^{-\tau-t})^2} \right] e^{-\delta t} dt \\
&= [3 - \Pi - (1 - \delta)(1 - e^{-\lambda})] \frac{1 + \delta}{\delta} \frac{1}{1 + \delta(1 - e^{-\tau})} - \frac{1 - e^{-\lambda}}{1 - e^{-\tau}} \\
&= \frac{1}{\delta[1 + \delta(1 - e^{-\tau})]} \left\{ (1 + \delta)[3 - \Pi - (1 - \delta)(1 - e^{-\lambda})] \right. \\
&\quad \left. - \delta[1 + \delta(1 - e^{-\tau})] \frac{1 - e^{-\lambda}}{1 - e^{-\tau}} \right\} \\
&= \frac{1}{\delta[1 + \delta(1 - e^{-\tau})]} \left\{ (1 + \delta)(2 - \Pi) + e^{-\lambda} - \delta \frac{1 - e^{-\lambda}}{1 - e^{-\tau}} \right\}.
\end{aligned}$$

Thus the expected discounted flow of future payoffs for $\tau \geq \lambda$ is given by

$$\begin{aligned}
\bar{v}(\tau, \lambda) &= \frac{1}{2\delta} [w_C(\underline{\varepsilon}) + w_C(\bar{\varepsilon})] \\
&\quad + \frac{(1 - e^{-\lambda})[2 + 3\delta + e^{-\lambda} - \delta \frac{1 - e^{-\lambda}}{1 - e^{-\tau}} - (1 + \delta)\Pi]}{4\delta[1 + \delta(1 - e^{-\tau})]} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})].
\end{aligned}$$

8.2 Proof of Proposition 1

The first-order condition of $\bar{v}(0, \lambda)$ with respect to λ is given by

$$\begin{aligned}
\frac{\partial \bar{v}(0, \lambda)}{\partial \lambda} &= \left\{ -\frac{\delta(\Pi - 1 - e^{-\lambda})e^{-(2+\delta)\lambda}}{4\delta[1 + \delta(1 - e^{-\lambda})]^2} \right. \\
&\quad \left. + \frac{(2 + \delta)e^{-\lambda} - (1 + \delta)(\Pi - 1)}{4\delta[1 + \delta(1 - e^{-\lambda})]} e^{-(1+\delta)\lambda} \right\} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})].
\end{aligned}$$

Hence,

$$\begin{aligned}
& \frac{\partial \bar{v}(0, \lambda)}{\partial \lambda} \geq 0 \\
& \Leftrightarrow -\delta(\Pi - 1 - e^{-\lambda})e^{-\lambda} \\
& \quad + [(2 + \delta)e^{-\lambda} - (1 + \delta)(\Pi - 1)][1 + \delta(1 - e^{-\lambda})] \geq 0 \\
& \Leftrightarrow (\Pi - 1)\{\delta e^{-\lambda} + (1 + \delta)[1 + \delta(1 - e^{-\lambda})]\} \\
& \quad \leq \delta e^{-2\lambda} + (2 + \delta)[1 + \delta(1 - e^{-\lambda})]e^{-\lambda} \\
& \Leftrightarrow \Pi \leq 1 + (1 + \delta) \frac{2 + \delta - \delta e^{-\lambda}}{(1 + \delta)^2 e^\lambda - \delta^2}. \tag{5}
\end{aligned}$$

Below, it is shown that the right-hand side of (5) unambiguously decreases in λ , taking on values in the range $[1, 2 + \frac{1}{1+2\delta}]$

$$\begin{aligned}
\frac{\partial(\frac{2+\delta-\delta e^{-\lambda}}{(1+\delta)^2 e^\lambda - \delta^2})}{\partial e^{-\lambda}} &= \frac{-\delta[(1 + \delta)^2 e^\lambda - \delta^2] + e^{2\lambda}(1 + \delta)^2[2 + \delta - \delta e^{-\lambda}]}{[(1 + \delta)^2 e^\lambda - \delta^2]^2} \\
&= \frac{\delta^3 + (1 + \delta)^2 e^\lambda [(2 + \delta)e^\lambda - 2\delta]}{[(1 + \delta)^2 - \delta^2 e^{-\lambda}]^2} > 0.
\end{aligned}$$

Since the right-hand side of (5) unambiguously decreases in λ , the following conclusion can be drawn: If $\Pi \leq 1$, then $\frac{\partial \bar{v}(0, \lambda)}{\partial \lambda} \geq 0$ for all $\lambda \geq 0$ and hence, $\hat{\lambda} = \infty$. If $\Pi \geq 2 + \frac{1}{1+2\delta}$, then $\frac{\partial \bar{v}(0, \lambda)}{\partial \lambda} \leq 0$ for all $\lambda \geq 0$ and hence, $\hat{\lambda} = 0$. If $\Pi \in (1, 2 + \frac{1}{1+2\delta})$, then there exists a unique $\hat{\lambda} \in (0, \infty)$ such that $\frac{\partial \bar{v}(0, \lambda)}{\partial \lambda} \geq 0$ if and only if $\lambda \leq \hat{\lambda}$, where $\hat{\lambda}$ is determined by equalizing the left-hand side with the right-hand side of condition (5)

$$\begin{aligned}
& \Pi = 1 + (1 + \delta) \frac{2 + \delta - \delta e^{-\hat{\lambda}}}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} \\
& \Leftrightarrow (1 + \delta)^2 (\Pi - 1) e^{\hat{\lambda}} - \delta^2 (\Pi - 1) - (1 + \delta)(2 + \delta) + \delta(1 + \delta) e^{-\hat{\lambda}} = 0 \\
& \Leftrightarrow e^{-2\hat{\lambda}} - \frac{\delta^2 (\Pi - 1) + (1 + \delta)(2 + \delta)}{\delta(1 + \delta)} e^{-\hat{\lambda}} + \frac{1 + \delta}{\delta} (\Pi - 1) = 0 \\
& \Leftrightarrow e^{-\hat{\lambda}} = \frac{2 + 3\delta + \Pi\delta^2 \pm \sqrt{(2 + 3\delta + \Pi\delta^2)^2 - 4\delta(1 + \delta)^3(\Pi - 1)}}{2\delta(1 + \delta)} \\
& \Leftrightarrow \hat{\lambda} = \ln \left[\frac{2\delta(1 + \delta)}{2 + 3\delta + \Pi\delta^2 + \sqrt{(2 + 3\delta + \Pi\delta^2)^2 - 4\delta(1 + \delta)^3(\Pi - 1)}} \right] \\
& \text{or } \hat{\lambda} = \ln \left[\frac{2\delta(1 + \delta)}{2 + 3\delta + \Pi\delta^2 - \sqrt{(2 + 3\delta + \Pi\delta^2)^2 - 4\delta(1 + \delta)^3(\Pi - 1)}} \right].
\end{aligned}$$

It is easily verified that the latter is the relevant solution (letting $\Pi \rightarrow 1$; it goes to infinity, while the former converges to $\ln(\frac{\delta}{2+\delta}) < 0$).

8.3 Proof of Lemma 1

Condition (5) from the previous proof can be rewritten as follows

$$\begin{aligned} \frac{\partial \bar{v}(0, \lambda)}{\partial \lambda} \geq 0 &\Leftrightarrow \Pi \leq 1 + (1 + \delta) \frac{2 + \delta - \delta e^{-\lambda}}{(1 + \delta)^2 e^{\lambda} - \delta^2} \\ &\Leftrightarrow \Pi \leq 1 + e^{-\lambda} \left[1 + \frac{(1 + \delta)e^{\lambda} - \delta}{(1 + \delta)^2 e^{\lambda} - \delta^2} \right]. \end{aligned} \quad (5')$$

It is the case that

$$\begin{aligned} \frac{\partial \left[\frac{(1+\delta)e^{\lambda} - \delta}{(1+\delta)^2 e^{\lambda} - \delta^2} \right]}{\partial \delta} &= \frac{[(1 + \delta)^2 e^{\lambda} - \delta^2](e^{\lambda} - 1) - 2[(1 + \delta)e^{\lambda} - \delta]^2}{[(1 + \delta)^2 e^{\lambda} - \delta^2]^2} \\ &= \frac{[(1 + 2\delta)e^{\lambda} + \delta^2(e^{\lambda} - 1)](e^{\lambda} - 1) - 2[e^{\lambda} + \delta(e^{\lambda} - 1)]^2}{[(1 + \delta)^2 e^{\lambda} - \delta^2]^2} \\ &= \frac{-\delta^2(e^{\lambda} - 1)^2 + e^{\lambda}[(1 - 2\delta)(e^{\lambda} - 1) - 2e^{\lambda}]}{[(1 + \delta)^2 e^{\lambda} - \delta^2]^2} \\ &= \frac{-\delta^2(e^{\lambda} - 1)^2 - e^{\lambda}[2 + (1 + 2\delta)(e^{\lambda} - 1)]}{[(1 + \delta)^2 e^{\lambda} - \delta^2]^2} < 0. \end{aligned}$$

Hence, the right-hand side of (5') unambiguously decreases in δ . Since the right-hand side of (5') unambiguously decreases in λ , it immediately follows that λ will have to fall for the right-hand side not to change as δ rises. Thus, a higher δ will be associated with a lower $\hat{\lambda}$. In particular, it follows that $\hat{\lambda}$ may become zero as δ is increased.

8.4 Proof of Lemma 2

By introspection of expression (1), it is easily seen that $\bar{w}(\tau, \hat{\lambda})$ is constant for $\tau \leq \hat{\lambda}$. For $\tau > \hat{\lambda}$, the first derivative is given by

$$\frac{\partial \bar{w}(\tau, \hat{\lambda})}{\partial \tau} = -\frac{e^{-\tau}(1 - e^{-\hat{\lambda}})[w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})]}{4(1 - e^{-\tau})^2} \left[3 - \Pi - 2\frac{1 - e^{-\hat{\lambda}}}{1 - e^{-\tau}} \right].$$

It is easy to see that the sign of $\frac{\partial \bar{w}(\tau, \hat{\lambda})}{\partial \tau}$ solely depends on the sign of $3 - \Pi - 2\frac{1 - e^{-\hat{\lambda}}}{1 - e^{-\tau}}$. Since $\Pi \in (1, 2 + \frac{1}{1+2\delta})$, it is straightforward that $3 - \Pi - 2\frac{1 - e^{-\hat{\lambda}}}{1 - e^{-\tau}} < 0$ and hence, $\frac{\partial \bar{w}(\tau, \hat{\lambda})}{\partial \tau} > 0$ for $\tau = \hat{\lambda}$. As τ increases,

$3 - \Pi - 2\frac{1-e^{-\hat{\lambda}}}{1-e^{-\tau}}$ increases monotonously, converging to

$$\begin{aligned} \lim_{\tau \rightarrow \infty} (3 - \Pi - 2\frac{1-e^{-\hat{\lambda}}}{1-e^{-\tau}}) &= 1 - \Pi + 2e^{-\hat{\lambda}} \\ &\stackrel{(5)}{=} 2e^{-\hat{\lambda}} - (1 + \delta) \frac{2 + \delta - \delta e^{-\hat{\lambda}}}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} \\ &= \frac{\delta(1 + \delta) + \delta(1 - \delta)e^{-\hat{\lambda}}}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} > 0. \end{aligned}$$

Hence, eventually $3 - \Pi - 2\frac{1-e^{-\hat{\lambda}}}{1-e^{-\tau}}$ will become positive and $\frac{\partial \bar{w}(\tau, \hat{\lambda})}{\partial \tau}$ will become negative. Thus, when $\hat{\lambda} \in (0, \infty)$ is implemented, \bar{w} initially increases and then decreases as τ increases beyond $\hat{\lambda}$ under the optimal solution. The value to which it will converge is calculated as follows

$$\begin{aligned} \lim_{\tau \rightarrow \infty} \bar{w}(\tau, \hat{\lambda}) &= \frac{1}{2}w_C(\underline{\varepsilon}) + \frac{1}{2}w_C(\bar{\varepsilon}) + \frac{1-e^{-\hat{\lambda}}}{4}(2 + e^{-\hat{\lambda}} - \Pi)[w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\ &= \frac{1}{4}[2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] + \frac{e^{-\hat{\lambda}}}{4}[\Pi - 1 - e^{-\hat{\lambda}}][w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\ &\stackrel{(5)}{=} \frac{1}{4}[2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] \\ &\quad + \frac{e^{-\hat{\lambda}}}{4} \left[\frac{(1 + \delta)(2 + \delta - \delta e^{-\hat{\lambda}})}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} - e^{-\hat{\lambda}} \right] [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\ &= \frac{1}{4}[2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] + \frac{e^{-\hat{\lambda}}[1 + \delta - \delta e^{-\hat{\lambda}}]}{4[(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2]} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\ &\tag{6} \\ &> \frac{1}{4}[2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] = \bar{w}(0, \hat{\lambda}). \end{aligned}$$

8.5 Proof of Lemma 3

First, consider the case when $\tau \leq \hat{\lambda}$. By introspection of expression (2), it is straightforward that \bar{v} increases if and only if $\Pi \geq 1 + e^{-\hat{\lambda}}$. For $\Pi \in (1, 2 + \frac{1}{1+2\delta})$, the optimal solution $\hat{\lambda}$ is implicitly defined by equalizing both sides of condition (5), which in the proof of lemma 1 was rewritten as (5')

$$\Pi = 1 + e^{-\hat{\lambda}} \left[1 + \frac{(1 + \delta)e^{\hat{\lambda}} - \delta}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} \right] > 1 + e^{-\hat{\lambda}}.$$

Hence, under the optimal solution, \bar{v} increases unambiguously for $\tau \leq \hat{\lambda}$. Next, consider the case when $\tau > \hat{\lambda}$. For $\tau > \hat{\lambda}$, the first derivative of \bar{v} with regard to τ is given by

$$\begin{aligned} & \frac{\partial \bar{v}(\tau, \hat{\lambda})}{\partial \tau} \\ &= \frac{e^{-\tau}(1 - e^{-\hat{\lambda}})[w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})]}{4[1 + \delta(1 - e^{-\tau})]^2} \left\{ [1 + \delta(1 - e^{-\tau})] \frac{1 - e^{-\hat{\lambda}}}{(1 - e^{-\tau})^2} \right. \\ & \quad \left. - [2 + 3\delta + e^{-\hat{\lambda}} - \delta \frac{1 - e^{-\hat{\lambda}}}{1 - e^{-\tau}} - (1 + \delta)\Pi] \right\} \\ &= \frac{e^{-\tau}(1 - e^{-\hat{\lambda}})[w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})]}{4[1 + \delta(1 - e^{-\tau})]^2} \left\{ (1 + \delta)(\Pi - 1) + 2\delta \left(\frac{1 - e^{-\hat{\lambda}}}{1 - e^{-\tau}} - 1 \right) \right. \\ & \quad \left. + \frac{1 - e^{-\hat{\lambda}}}{(1 - e^{-\tau})^2} - (1 + e^{-\hat{\lambda}}) \right\}. \end{aligned}$$

Hence,

$$\frac{\partial \bar{v}(\tau, \hat{\lambda})}{\partial \tau} \geq 0 \Leftrightarrow (1 + \delta)(\Pi - 1) + 2\delta \left(\frac{1 - e^{-\hat{\lambda}}}{1 - e^{-\tau}} - 1 \right) + \frac{1 - e^{-\hat{\lambda}}}{(1 - e^{-\tau})^2} - (1 + e^{-\hat{\lambda}}) \geq 0.$$

It is easy to see that the left-hand side decreases monotonously in τ . It equals $(1 + \delta)(\Pi - 1) + \frac{e^{-2\hat{\lambda}}}{1 - e^{-\hat{\lambda}}} > 0$ for $\tau = \hat{\lambda}$. Hence, \bar{v} initially increases as τ increases beyond $\hat{\lambda}$. The left-hand side decreases monotonously in τ , converging to

$$(1 + \delta)(\Pi - 1 - 2e^{-\hat{\lambda}}) \stackrel{(5)}{=} \frac{1 + \delta}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} [-\delta(1 + \delta) - \delta(1 - \delta)e^{-\hat{\lambda}}] < 0.$$

Hence, \bar{v} will initially increase and eventually decrease. The value, to which $\bar{v}(\tau, \hat{\lambda})$ converges is obtained by appropriately discounting the value to which \bar{w} converges, given by (6):

$$\begin{aligned} & \lim_{\tau \rightarrow \infty} \bar{v}(\tau, \hat{\lambda}) \\ &= \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] + \frac{e^{-\hat{\lambda}}[1 + \delta - \delta e^{-\hat{\lambda}}]}{4\delta[(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2]} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})]. \end{aligned}$$

Using the optimality condition for $\hat{\lambda}$, given by equalizing both sides of condition (5), the expression for $\bar{v}(0, \hat{\lambda})$, given by (2), can be rearranged

as follows

$$\begin{aligned}
\bar{v}(0, \hat{\lambda}) &= \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] \\
&\quad + \frac{[\Pi - (1 + e^{-\hat{\lambda}})]e^{-(1+\delta)\hat{\lambda}}}{4\delta[1 + \delta(1 - e^{-\hat{\lambda}})]} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\
&\stackrel{(5)}{=} \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] \\
&\quad + \frac{e^{-(1+\delta)\hat{\lambda}}}{4\delta[1 + \delta(1 - e^{-\hat{\lambda}})]} \left[\frac{(1 + \delta)(2 + \delta) - \delta(1 + \delta)e^{-\hat{\lambda}}}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} - e^{-\hat{\lambda}} \right] [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})] \\
&= \frac{1}{4\delta} [2w_N(\underline{\varepsilon}) + w_N(\bar{\varepsilon}) + w_C(\bar{\varepsilon})] + \frac{e^{-(1+\delta)\hat{\lambda}}}{4\delta[(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2]} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})].
\end{aligned}$$

Since $1 + \delta - \delta e^{-\hat{\lambda}} > 1 > e^{-\delta\hat{\lambda}}$ for $\hat{\lambda} > 0$, it immediately follows that $\lim_{\tau \rightarrow \infty} \bar{v}(\tau, \hat{\lambda}) > \bar{v}(0, \hat{\lambda})$.

8.6 Proof of Proposition 3

For $\tau > \hat{\lambda}$, the first derivative with regard to λ is given by

$$\frac{\partial \bar{v}(\tau, \lambda)}{\partial \lambda} = \frac{e^{-\lambda} [w_N(\underline{\varepsilon}) - w_C(\underline{\varepsilon})]}{4\delta[1 + \delta(1 - e^{-\tau})]} \left[(1 + \delta)(1 - \Pi) + 2e^{-\lambda} + 2\delta \left(1 - \frac{1 - e^{-\lambda}}{1 - e^{-\tau}} \right) \right].$$

Define $D(\tau) \equiv (1 + \delta)(1 - \Pi) + 2e^{-\lambda} + 2\delta \left(1 - \frac{1 - e^{-\lambda}}{1 - e^{-\tau}} \right)$. It is straightforward that $\frac{\partial \bar{v}(\tau, \lambda)}{\partial \lambda} \geq 0$ if and only if $D(\tau) \geq 0$. It is easy to see that D increases in τ and decreases in λ . Since $D(\tau) > 0$ for $\lambda = 0$ and $D(\tau) < 0$ as $\lambda \rightarrow \infty$, it immediately follows that for any given τ , there exists a unique $\lambda \in (0, \infty)$ such that $D(\tau) = 0$ and hence $\bar{v}(\tau, \lambda)$ is maximized. Moreover, it follows that the solution to $\frac{\partial \bar{v}(\tau, \lambda)}{\partial \lambda} = 0$ increases in τ . Hence, the ex post optimal solution will increase over time.

It remains to establish the relation of the ex post optimal adjustment phase length to the one being employed in the agreement. Setting $\lambda = \hat{\lambda}$, the value for D when $\tau = \hat{\lambda}$ is given by

$$\begin{aligned}
D(\hat{\lambda}) &= (1 + \delta)(1 - \Pi) + 2e^{-\hat{\lambda}} \\
&\stackrel{(5)}{=} 2e^{-\hat{\lambda}} - (1 + \delta)^2 \frac{2 + \delta - \delta e^{-\hat{\lambda}}}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} \\
&= \frac{-\delta(1 + \delta)^2(1 - e^{-\hat{\lambda}}) - 2\delta^2 e^{-\hat{\lambda}}}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} < 0.
\end{aligned}$$

Hence, the ex post optimal solution for the time limit will eventually be smaller than the one agreed-upon as τ increases beyond $\hat{\lambda}$. The value for D as $\tau \rightarrow \infty$ when $\lambda = \hat{\lambda}$ is given by

$$\begin{aligned} \lim_{\tau \rightarrow \infty} D(\tau) &= (1 + \delta)(1 - \Pi + 2e^{-\hat{\lambda}}) \\ &\stackrel{(5)}{=} (1 + \delta) \left[2e^{-\hat{\lambda}} - (1 + \delta) \frac{2 + \delta - \delta e^{-\hat{\lambda}}}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} \right] \\ &= (1 + \delta) \frac{\delta(1 + \delta) + \delta(1 - \delta)e^{-\hat{\lambda}}}{(1 + \delta)^2 e^{\hat{\lambda}} - \delta^2} > 0. \end{aligned}$$

Hence, the ex post optimal solution for the time limit will eventually be larger than the one prescribed by the agreement.