A Political-Economy Theory of Trade Agreements*

Giovanni Maggi
Princeton University and NBER

Andres Rodríguez-Clare
IADB

June 2005

Abstract

We develop a model of trade agreements where governments are motivated by the desire to commit vis-a-vis domestic industrial lobbies. In contrast to the standard theory, where trade agreements are solely designed to deal with terms-of-trade externalities, here "politics" affect the extent of trade liberalization that takes place in the agreement. In particular, trade agreements entail deeper trade liberalization when governments are more politically motivated (i.e. care less about social welfare and more about contributions), when governments have less bargaining power vis-á-vis domestic lobbies, and when lobbies have less influence on the negotiation of the agreement. An interesting implication of the model is that, under certain conditions, governments prefer to write "incomplete" agreements that only specify maximum tariffs, rather than exact tariff levels. We also find that trade liberalization is deeper when resources are more mobile across sectors. When we consider a continuous-time extension of the model, we find that the optimal agreement is made of two components: an immediate slashing of tariffs relative to their noncooperative levels, and a subsequent, gradual reduction of tariffs. The instantaneous tariff cut is a reflection of the terms-of-trade motive for the agreement, while the domestic-commitment motive is reflected in the gradual part of trade liberalization. In the gradual phase, the speed of liberalization is higher when resources are more mobile.

*We thank Paula Auerbach for helping us draw the figures and seminar participants at ECARES, Ente Luigi Einaudi di Roma, Federal Reserve Bank of New York, FGV Rio de Janeiro, IADB, Penn State University, Southern Methodist University, Syracuse University, University of Texas at Austin, University of Virginia, and Vanderbilt University for very useful comments. Giovanni Maggi thanks the National Science Foundation for financial support.

1. Introduction

The history of trade liberalization after World War II is intimately related with the creation and expansion of the GATT (now WTO), and with the signing of countless bilateral and regional trade agreements. Clearly, there are strong forces pushing countries to sign international trade agreements, and it is important for economists and political scientists to understand what these forces are. Why do countries engage in trade agreements? What determines the extent and form of liberalization that takes place in such agreements?

The standard theory of trade agreements dates back to Johnson (1954), who argued that, in the absence of trade agreements, countries would attempt to exploit their international market power by taxing trade, and the resulting equilibrium (trade war) would be inefficient for all countries involved. International trade agreements can be seen as a way to prevent such a trade war. This idea was later formalized in modern game-theoretic terms by Mayer (1981).

Grossman and Helpman (1995a) and Bagwell and Staiger (1999) have extended this framework to settings where governments are subject to political pressures. In their models, even politically-motivated governments engage in trade agreements *only* to correct for terms of trade externalities. Thus, "politics" do not affect the motivation to engage in trade agreement nor the extent of liberalization that they bring about.

In this paper we present a theory where politics is very much at the center of trade agreements. In particular, we consider a model where trade agreements help governments to deal with a time-inconsistency problem in their interaction with domestic lobbies. In a previous paper (Maggi and Rodríguez-Clare, 1998) we showed how such a time-inconsistency problem may emerge when capital is fixed in the short run but mobile in the long run, and argued that this may motivate even a small economy to seek commitment through a trade agreement. The present paper builds on this idea to develop a fuller theory of trade agreements.

We start by reviewing the logic behind the domestic-commitment problem that is at the basis of our theory. This logic is easily illustrated for the case of a small open economy. According to the modern political-economy theory of trade policy, it is not clear why a small-country government would want to "tie its hands" and give up its ability to grant protection. For example, in Grossman and Helpman (1994), lobbies compensate the government for the distortions associated with trade policy, and hence there is no reason why the government would want to commit not to grant protection. In fact, if the government is able to extract rents from the political process it is strictly better off in the political equilibrium than under free trade. But this is no longer necessarily true when one takes into account capital mobility. This is because, given the expectation of protection in one sector, there may be excessive investment there. Since this happens before the government and lobbies sit down to negotiate over protection, the government is not compensated for this "long-run" distortion. If this distortion dominates the government's short-run gains from protection, the government would gain by committing to free trade.

Our previous paper formalizes the basic idea just described, but falls short of a full theory of trade agreements, because of three limitations. First, that model considers only a small economy. A proper theory of trade agreements should consider at least two countries. For example, with a small country model it is impossible to study the interaction between the terms-of-trade motive and the domestic-commitment motive for a trade agreement. Second, in that paper the government is only allowed to choose between two extreme options, namely free trade or no commitment at all. If we want to study what determines the *extent* of trade liberalization, we need to allow governments to commit to intermediate levels of trade protection. Third, our previous paper allows lobbies to interact with the government only after the agreement has been signed ("ex-post" lobbying), but does not allow lobbies to influence the selection of the agreement ("ex-ante" lobbying). Ex-ante lobbying is arguably important in reality. How are the results affected by this? Does ex-ante lobbying eliminate the governments' drive for liberalization through trade agreements?

The present paper analyzes trade agreements in a model where two large countries can commit to arbitrary trade taxes, and lobbies may be active both at the ex-ante stage and at the ex-post stage. In addition, the model allows for imperfect mobility of capital across sectors. An attractive feature of our model is that it integrates both existing motives for trade agreements, namely terms-of-trade externalities and domestic-commitment problems. Moreover, the model leads to several implications that appear consistent with casual empirical observations.

First, in contrast to the "standard" theory of trade agreements, here the extent of trade liberalization is affected by politics. If governments are more politically motivated, in the sense that they care less about social welfare and more about political contributions from interest groups, then trade agreements will bring about more substantial trade liberalization. Moreover, trade liberalization is less deep when governments have less bargaining power vis-á-vis domestic lobbies, and when lobbies have less influence on the negotiation of the agreement.

Second, our model can explain why trade agreements typically specify *maximum* tariff levels ("tariff ceilings") rather than exact tariff commitments ("exact tariffs"). In other words, the model can explain why agreements are incomplete contracts, and in particular why they leave a certain amount of discretion to governments. Tariff ceilings and exact tariffs have very different implications. With exact tariffs, lobbying effectively ends at the time of the agreement, since the agreement leaves no discretion for governments to choose tariffs in the future. With tariff ceilings, on the other hand, governments retain the option of setting tariffs below their maximum levels, and this will invite lobbying also after the agreement is signed. Whether or not the agreement will induce ex-post lobbying depends crucially on the strength of ex-ante lobbying. We find that, if ex-ante lobbying is not too strong, tariff ceilings tend to be preferred to exact tariff commitments.[1]

Third, in our model the degree of capital mobility is a key determinant of the extent of trade liberalization. We find that trade liberalization is deeper when capital is more mobile across sectors. To understand this result, consider the extreme case in which capital can be freely reallocated after the agreement has been signed. In this case, the lobby suffers no loss from trade liberalization, since capital can exit the affected sector and avoid any losses associated with lower domestic prices. With imperfect capital mobility, however, liberalization does generate losses for the lobby, and as a consequence ex-ante lobbying is stronger. Although our model generates only a comparative-statics result, it nevertheless suggests a cross-sectional

[1] Ours is not the only possible explanation for the use of tariff ceilings. Bagwell and Staiger (2005) propose a model where tariff ceilings are motivated by the presence of privately observed shocks in the political pressures faced by governments.

empirical prediction: we should observe deeper trade liberalization in sectors where capital is more mobile. We are not aware of any empirical work exploring the link between factor mobility and trade liberalization, but casual observations seem to be in line with our model's prediction: for example, trade liberalization has been very limited in the agricultural sector, which is intensive in resources that are not very mobile (e.g. land).

Fourth, when we extend the model to a continuous-time setting, we find that the optimal agreement is made of two components: an immediate slashing of tariffs relative to their non-cooperative levels, and a subsequent, gradual reduction of tariffs. The initial tariff reduction captures the terms-of-trade motive for the trade agreement, while the domestic-commitment motive is reflected in the gradual component of trade liberalization. We also find that the *speed* of trade liberalization is higher when capital is more mobile. While our model is not the only one that can explain gradual trade liberalization, the explanation proposed here, based on domestic commitment problems and imperfect capital mobility, is novel and – we feel – empirically plausible.

We want to emphasize that most of our insights follow from our structural modeling of the lobbying game, in which interest groups and governments exchange contributions for trade protection. If we modeled political pressures with a reduced-form approach, keeping lobbies and contributions in the background, we would lose most of our results. For example, one might be tempted to model the domestic-commitment problem by assuming that there is a divergence between ex-ante and ex-post government objectives (e.g. at the stage of signing the agreement governments maximize welfare, while ex-post they maximize a combination of welfare and industry profits). This reduced-form setup would *not* be equivalent to our structural setup: for example, in the reduced-form setup there would be no role for tariff ceiling, and there would be no gradualism in trade liberalization.

This paper is related to two literatures: first, the literature on trade agreements motivated by terms-of-trade externalities (see the papers cited at the beginning of this introduction); and second, the literature on trade agreements motivated by domestic-commitment problems. In this second group, Maggi and Rodríguez-Clare (1998) and Mitra (2002) have highlighted the role of politics in creating demand for commitment, while Staiger and Tabellini (1987) have highlighted purely economic considerations. However, these three papers focus on a single small economy and do not attempt a full-fledged analysis of trade agreements.

A recent paper that considers a full model of trade agreements in the presence of domestic commitment problems is Conconi and Perroni (2005). They consider a self-enforcing agreement between a large country and a small country, where the only motive for a trade agreement is a domestic commitment issue that affects the small country.[2] In contrast, our model integrates both motives for trade agreements, namely terms-of-trade externalities and domestic-commitment problems. Another important difference is that they take a reduced-form approach where there is a divergence between ex-ante and ex-post objectives of the governments. As we

[2] Conconi and Perroni (2004) consider a model of self-enforcing international agreements between two large countries where there is both a domestic commitment problem and an international externality. This paper is different from ours in that it analyzes issues of self-enforcement in a model with very little structure and thus offers no implications for the extent of trade liberalization brought about by trade agreements, which is the focus of our present paper.

pointed out above, this approach is not equivalent to our structural approach where lobbying and contributions are explicitly modeled; most of our points could not be made with a reduced-form approach. In any event, Conconi and Perroni's paper makes very different points from ours, as they focus on the implications of the self-enforcement constraints and argue that they can explain the granting of temporary Special and Differential treatment to developing countries in the WTO.

Another literature that is related to our paper is that on gradual trade liberalization. In most of these papers, e.g. Staiger (1995), Furusawa and Lai (1999), Bond and Park (2004) and Conconi and Perroni (2005), gradual trade liberalization is a consequence of the self-enforcing nature of the agreements. In these models trade liberalization occurs at once if players are sufficiently patient. The explanation we propose in the present paper does not rely on self-enforcement considerations, but rather on the interaction between frictions in capital mobility and lobbying by capital owners.

The paper is organized as follows. The next section presents the basic two-period model. Section 3 extends the model to a continuous time setting in order to explore gradual trade liberalization. Section 3 considers some extensions of the model. Section 4 concludes the paper.

2. The model

2.1. The economic structure

There are two countries, Home and Foreign, and three goods: one numeraire good, denoted by N, and two manufacturing goods, denoted by M_1 and M_2.

There are two types of capital, type 1 and type 2. The M_1 good is produced one-for-one from type-1 capital, the M_2 good is produced one-for-one from type-2 capital. Each country is endowed with one unit of each type of capital. The only difference between the two countries is in the technology to produce the N good: in country H, the N good is produced one-for-one from type-1 capital, while in country F, the N good is produced one-for-one from type-2 capital. Given these assumptions, under free trade Home exports good M_2 and Foreign exports good M_1. The reason we chose this particular technology structure is that it generates a simple symmetric setup where, in each country, capital mobility is relevant only between the import-competing sector and the numeraire sector. This in turn ensures that in each country the domestic-commitment motive for trade agreements concerns the import-competing sector but not the export sector, a feature that simplifies the analysis considerably.

In both countries preferences are given by

$$U = c_N + \sum_{i=1}^{2} u(c_i)$$

where $u(c_i) = vc_i - c_i^2/2$. Thus the demand function for good M_i is

$$d(p_i) = v - p_i$$

Home chooses a specific tariff t on M_1 and Foreign chooses a specific tariff t^* on M_2. Thus, if tariffs are not prohibitive, the domestic price of good M_1 in Home is given by $p_1 = p_1^* + t$. Similarly, the domestic price of good M_2 in Foreign is $p_2^* = p_2 + t^*$.[3]

Let x (x^*) denote the level of capital allocated to sector M_1 (M_2) in country H (F). Welfare (i.e., utility of the representative agent) is given by income plus rents of capitalists plus tariff revenue plus consumer surplus from manufacturing goods. Thus, welfare in Home and Foreign, respectively, is given by:

$$\begin{aligned} W &= (1-x) + (p_1 x + t m_1 + s_1) + (p_2 + s_2) \\ W^* &= (1-x^*) + (p_2^* x^* + t^* m_2^* + s_2^*) + (p_1^* + s_1^*) \end{aligned}$$

Note the separability between sectors M_1 and M_2. Specifically, note that we can express W as the sum of two components: the first one, $(1-x) + (p_1 x + t m_1 + s_1)$, depends on t and x; and the second one, $p_2 + s_2$, depends on t^* and x^*. The same separability applies to foreign welfare. Together with symmetry, this separability implies that we can focus on sector M_1; the equilibrium in M_2 will be a mirror image. Thus, to simplify notation, we drop the subscript 1 from now on, and simply refer to good M_1 as "manufacturing" or "manufactures."

The international market clearing condition for manufacturing is:

$$d(p) + d(p^*) = x + 1$$

This yields

$$p^*(t,x) = v - \frac{1}{2}(x+1+t)$$

and

$$p(t,x) = v - \frac{1}{2}(x+1-t)$$

where we emphasize the dependence of equilibrium prices on the tariff and the capital allocation in the home country.

Letting $m = d(p) - x$ denote imports of manufactures by Home, then:

$$m(t,x) = \frac{1}{2}(\Delta x - t)$$

where $\Delta x \equiv 1 - x$ is the difference in supply between the two countries.

Given this notation, welfare in Home is:

$$W(t,x) = (1-x) + p(t,x)x + tm(t,x) + s(t,x) + [\cdot]$$

[3] In this paper we do not consider export subsidies and taxes. If the agreement takes the traditional form of exact tariff and subsidy commitments, this restriction is innocuous, because only net protection (i.e. the difference between import tariff and export subsidy in a given sector) matters, therefore t and t^* can be reinterpreted in terms of net protection in the two sectors. If the agreement takes the form of tariff and subsidy ceilings, on the other hand, not only net protection but also the levels of import tariffs and export subsidies matter, and this makes the analysis substantially more complex. In Maggi and Rodriguez-Clare (2005) we study optimal agreements when both import and export instruments are allowed but there is no capital mobility. We also note that assuming away export instruments is relatively common in the existing literature on trade agreements: see for example Grossman and Helpman (1995b), Krishna (1998), Maggi (1999) and Ornelas (2004).

where $[\cdot]$ does not depend on t and x. Analogously, Foreign welfare is

$$W^*(t,x) = p^*(t,x) + s^*(t,x) + [\cdot]$$

At this point it is useful to derive the free trade equilibrium. Under free trade and perfect capital mobility, the domestic (and international) price of the M_1 good must be equal to one. Thus the free trade allocation of capital, x^{ft}, is determined by

$$v - \frac{1}{2}(x^{ft} + 1) = 1$$

To ensure that under free trade Home is incompletely specialized and imports good M_1 we need $0 < x^{ft} < 1$, which holds as long as $3/2 < v < 2$. We maintain this assumption throughout the rest of the paper.

Note that, because of the symmetry of the model, under free trade there is no trade in the numeraire sector.

2.2. The political structure

We assume that, in each country, the capital owners in the import-competing sector get organized as a lobby and offer contributions to their government in exchange for protection.[4]

We model the interaction between lobby and government in a similar way as Grossman and Helpman (1994). The government's objective function is

$$aW + C$$

where C denotes contributions from the import-competing lobby. The parameter a captures (inversely) the importance of political considerations in the government's objective: when a is lower, "politics" are more important.

The lobby maximizes total returns to capital net of contributions:[5]

$$px - C$$

The lobby collects contributions in proportion to the amount of capital, thus total contributions are given by $C = cx$, where c is the contribution per unit of capital.

The timing of the non-cooperative game is the following. In the first stage, investors allocate their capital. The value of x summarizes the choices of investors in the first stage. In the second stage, the government and the import competing lobby in each country bargain efficiently over tariff and contributions. For simplicity we assume that the lobby has all the bargaining power (we relax this assumption in a later section). An equivalent assumption would be that the lobby makes a take-it-or-leave-it offer to the government that consists in a tariff level and a contribution level.

[4] We are implicitly assuming that the export sector and the numeraire sector are not able to get organized. This is a simple lobby structure that generates trade protection in the political equilibrium.

[5] This is a shortcut. To be more precise, we should specify the lobby's objective as the aggregate well-being of its lobby members, but this would give rise to the same results. First recall that, given separability and no export policies, we can take a partial equilibrium approach in writing the lobby's objective. Letting α be the fraction of the population that owns some capital in the import-competing sector, the lobby's objective is $px + \alpha(tm + s) - C$, so the joint surplus of government and lobby is proportional to $\frac{a+\alpha}{1-\alpha}W + px$, an expression that has the same qualitative structure as the one we derive below.

2.3. The short-run noncooperative equilibrium

To determine the subgame perfect equilibria of the game we proceed by backward induction, starting with the determination of equilibrium tariffs and contributions given the allocation of capital. This is the equilibrium of the subgame, or the "short-run" equilibrium.

We can focus on the Home country. Given x, the government (G) and the lobby (L) choose t to maximize

$$J^{SR}(t,x) = aW(t,x) + p(t,x)x$$

This yields

$$t = t^J(x) \equiv (1/3)(\Delta x + 2x/a)$$

The noncooperative tariff t^J can be decomposed in two parts. The component $\Delta x/3$ captures the incentive to distort terms of trade: when the supply difference Δx is bigger, the volume of imports is larger, and hence this incentive is stronger. The component $2x/3a$ captures the political influence exerted by the lobby. This component is more important when the sector is larger (x is higher) and when the government's valuation of contributions relative to welfare is higher (a is lower). For future reference, note that as $a \to \infty$,

$$t \to t^W(x) \equiv \Delta x/3$$

This is simply the standard welfare-maximizing tariff for a large country given the basic economic structure above.

For future reference, we define $c(t,x)$ as the contributions per unit of capital such that G is just willing to impose tariff t, or in other words, such that G is kept at its reservation utility given tariff t. In the absence of contributions, G would choose the welfare maximizing tariff given x, that is $t^W(x)$, so G's reservation utility is $W(t^W(x),x)$. Since the short-run equilibrium tariff given x cannot be below $t^W(x)$, we only need to focus on the case $t \geq t^W(x)$. For the government to choose a tariff $t \geq t^W(x)$, total contributions would have to be equal to

$$\begin{aligned} a\left[W(t^W(x),x) - W(t,x)\right] &= -\int_{t^W(x)}^{t} aW_1(t,x)dt \\ &= -\int_{\Delta x/3}^{t} (a/4)(\Delta x - 3t)dt \\ &= (3a/8)\left(t - t^W(x)\right)^2 \end{aligned}$$

Thus, the function

$$c(t,x) \equiv (3a/8x)\left(t - t^W(x)\right)^2$$

determines the contributions per unit of capital necessary to induce the government to choose tariff $t \geq t^W(x)$. Note that this function is defined only for $t \geq t^W(x)$, since the lobby would never be willing to pay the government to impose a tariff lower than it would choose on its own.

2.4. The long-run noncooperative equilibrium

We now turn to the long-run non-cooperative equilibrium, where x is endogenous and is determined according to investors' expectations about future protection in the absence of a trade agreement.

The equilibrium conditions are then:

$$t = t^J(x)$$
$$p(t, x) - c(t, x) = 1 \tag{2.1}$$

The second condition requires that the return to capital net of contributions be equal in the import-competing sector and in the numeraire sector.[6] Thia equal-returns condition implicitly defines a curve in (t, x) space, that we label $x^{er}(t)$ (we will sometimes use $t^{er}(x)$ for its inverse). Note that, since we defined the function $c(t, x)$ only for $t \geq t^W(x)$, the curve $x^{er}(t)$ is defined only in the region $t \geq t^W(x)$. We let $(\hat{t}, \hat{x})$ denote a solution to the above system. Also, we let $\left(t^W, x^W\right)$ denote the point at the intersection of the curves $t^W(x)$ and $x^{er}(t)$.

Proposition 1. *If $a > (6v - 7)/6(2 - v)$ there exists a unique long-run noncooperative equilibrium. In this equilibrium each country imposes a positive but non-prohibitive tariff $\hat{t}$. The equilibrium tariff $\hat{t}$ is decreasing in a, and approaches t^W as $a \to \infty$.*

Proof: see Appendix.

Figure 1 illustrates the long-run noncooperative equilibrium. In the figure, the $t^J(x)$ curve is increasing, but nothing would change if it were decreasing. Under the condition assumed in Proposition 1, curve $x^{er}(t)$ has a unique intersection with the $t^J(x)$ curve, and at this intersection it has an infinite slope. Also, $x^{er}(t)$ is increasing in the region between the curves $t^W(x)$ and $t^J(x)$ (see proof of Proposition 1). We will maintain the assumption $a > (6v - 7)/6(2 - v)$ throughout the paper.

Not surprisingly, for any positive but finite level of a, the non-cooperative equilibrium described in Proposition 1 entails a tariff higher than the one that would prevail in the case of "no politics" (i.e., $a \to \infty$). In other words, $\hat{t} > t^W$. This implies that $\hat{x} > x^W$. The reason for this is that $t^J(x^W) > t^W$, implying that

$$p(t^J(x^W), x^W) - c(t^J(x^W), x^W) > 1$$

Hence, if $x = x^W$ the associated short-run equilibrium leads to higher net returns in manufacturing than in the numeraire sector, which in turn leads to a higher x. As we will show formally

[6] An alternative way to find the long-run equilibrium, perhaps more standard from a game theoretical point of view, would be to derive the equilibrium of the subgame given x and then proceed by backward induction to derive the equilibrium level of x. More specifically, given x, the subgame equilibrium contribution and tariff are respectively given by $c(x) = (3a/8x)\left(t^J(x) - t^W(x)\right)^2$ and $t(x) = t^J(x)$. The equilibrium x can then be found as the one that solves $p(t^J(x), x) - c(x) = 1$. The procedure we follow in the text turns out to be more convenient for the analysis of the optimal trade agreement.

in a later section, this excess of $\hat{x}$ above x^W represents an overinvestment problem, that is a "long-run" distortion associated with the Government's lack of commitment vis-á-vis domestic investors. Each government is compensated by its lobby for the short-run distortion associated with protection (i.e. the consumption distortion given the allocation x), but is not compensated for the long-run distortion. For this reason a government may value a commitment to a lower level of the tariff. This is the heart of the domestic-commitment motive for trade agreements, which operates alongside the standard terms-of-trade motive. We are now ready to examine the optimal agreement.

2.5. The optimal trade agreement

We suppose that, before capital is allocated, the two governments and the two lobbies determine the trade agreement. In Maggi and Rodríguez-Clare (1998) we assumed that lobbies do not influence the selection of the trade agreement, i.e. there is no ex-ante lobbying. Here we allow for ex-ante lobbying by assuming that the agreement maximizes the ex-ante joint surplus of the two governments and the two lobbies.[7] To capture the strength of ex-ante lobbying, we weigh the lobbies' part of the joint surplus with a parameter δ. The agreement maximizes the following objective:

$$\Psi = U^G + U^{G^*} + \delta(U^L + U^{L^*})$$

where U^G, U^{G^*}, U^L and U^{L^*} denote the second-period payoffs of the governments and lobbies as viewed from the ex-ante stage. We will be more explicit about these payoffs shortly, but first we want to discuss the interpretation of δ.

A lower level of δ is interpreted as a situation where lobbies have less influence on the shaping of the ex-ante agreement. The case $\delta = 1$ corresponds to the benchmark case in which ex-ante lobbying is just as strong as ex-post lobbying. The case $\delta = 0$ corresponds to the case in which there is no ex-ante lobbying, as we assumed in our previous paper. We can offer two justifications of δ in terms of more fundamental parameters:

1. A direct interpretation of δ would be as the discount factor of the lobbies relative to that of the governments. We have in mind that an agreement is a long-run commitment that shapes the political game for a long time to come, therefore discounting considerations are potentially important for the determination of the agreement. The discount factor of a government can be thought of as determined by political factors such as the probability of re-election, and the discount factor of a lobby is determined by economic factors such as the probability of bankruptcy.

2. An alternative story for δ is the following. Suppose that governments and lobbies discount the future in the same way, but contributions may be more effective in influencing day-to-day

[7] This efficient-agreement approach can be justified as equivalent to a more structural game between governments and lobbies. One possibility would be to consider a game along the lines of Grossman and Helpman (1995). They assume that lobbies offer (differentiable) contribution schedules to their respective governments and then governments bargain efficiently given the contribution schedules, and show that the equilibrium outcome maximizes the joint surplus of governments and lobbies. More generally, any negotiation procedure between governments and lobbies that yields a joint-surplus-maximizing outcome would be equivalent to our approach.

policy decisions than they are in influencing the negotiation of the agreement.[8] To capture this idea we can write the home government's ex-ante objective as $U^G_{ex-ante} = U^G + \delta C_{ex-ante}$, with an analogous expression holding for the foreign government. Here δ captures the weight of ex-ante contributions, which can be different from that of ex-post contributions. We can write the domestic lobby's ex-ante objective as $U^L_{ex-ante} = U^L - C_{ex-ante}$, and analogously for the foreign lobby. Multiplying the lobbies' payoffs by δ to make utility transferable ex-ante, and summing up, the ex-ante joint surplus can then be written as $\Psi = U^G + U^{G^*} + \delta(U^L + U^{L^*})$.

In principle, ex-ante lobbying might be stronger than ex-post lobbying ($\delta > 1$), for example if governments are more shortsighted than lobbies. For this reason we will allow δ to be higher than one.

Agreements are assumed to be perfectly enforceable. In the concluding section we will discuss how the insights of our model might extend to a setting of self-enforcing agreements.

We assume that the inherited level of x at the agreement stage is equal to $\widehat{x}$, the long-run equilibrium allocation in the absence of an agreement. The interpretation is that the commitment opportunity comes as a surprise to the private sector.

Following the agreement, each capital owner gets a chance to move its unit of capital with probability $z \in [0, 1]$. Thus, a fraction z of the capital has the opportunity to move. The case $z = 0$ captures the case of fixed capital, whereas the case $z = 1$ captures a situation in which capital is perfectly mobile in the long run but fixed in the short run. With a slight abuse of terminology, from now on we refer to this case simply as perfect capital mobility, and to the case $z < 1$ as imperfect capital mobility.

To recapitulate, the timing of the model is as follows:

1. The agreement is selected;
2. Capital is reallocated (when feasible);
3. Given the capital allocation and the constraints (if any) imposed by the agreement, each government-lobby pair chooses a tariff.

Again, given separability across the two manufacturing sectors, we can analyze them independently, knowing that the agreement for sector M_2 will be the mirror image of that for M_1. Moreover, given symmetry, the optimal agreement must maximize the joint surplus of the two governments and the import-competing lobby in each sector. Thus, just as in the previous subsections, we can focus on sector M_1 (omitting subscripts) and find the optimal agreement by maximizing the joint surplus of the two governments and Home's import-competing lobby in this sector.

There are two forms of agreement that we can consider: agreements that specify tariff ceilings, that is constraints of the type $t \leq \bar{t}$, and agreements that specify exact tariffs, that is constraints of the type $t = \bar{t}$. The main difference between these two types of agreement is that

[8] For example, it might be the case that the US trade representatives involved in the negotiation of trade agreements attach less value to contributions than policymakers involved in day-to-day decisions on trade policy (e.g. members of Congress). This is plausible if contributions are used to finance electoral campaigns, since trade negotiators are not elected officials. Another possibility is that, since a trade agreement is a long-run commitment, the magnitude of contributions required to influence it is higher than that required to influence day-to-day policy choices, and there may be a political cost associated with paying larger contributions because they are more visible to voters.

in the case of exact tariff commitments the lobby will not have to pay contributions to obtain protection ex-post, since such protection will effectively be part of the agreement. Under tariff ceilings, on the other hand, the government can credibly threaten to impose its unilateral best tariff $t^W(x)$. Thus, the lobby would have to compensate the government for deviating from this tariff, and there would be positive contributions ex-post. Clearly, whether tariff ceilings or exact tariffs are preferred crucially depends on the value of δ, since this determines how ex-ante joint surplus depends on ex-post contributions. The following proposition establishes conditions on δ under which there is no loss of generality in focusing on tariff ceilings:

Proposition 2. *There exists a $\bar{\delta} \geq 1$ such that, if $\delta \leq \bar{\delta}$, tariff ceilings do at least as well as exact tariff commitments. If capital is perfectly mobile, this is true for any δ.*

Proof: See Appendix.

This result states that, if ex-ante lobbying is not too strong, there is no loss of generality in focusing on tariff ceilings. The intuition is simple. Consider first the case of perfect capital mobility. Tariff ceilings are preferable to exact tariffs for two reasons: (1) If the agreement imposes exact tariffs, clearly there will be no ex-post lobbying, and hence no ex-post contributions. On the other hand, tariff ceilings may induce ex-post contributions: if the ceiling for the tariff is sufficiently high, the lobby will offer contributions to convince the government to raise the tariff towards the ceiling. From the point of view of the ex-ante joint surplus, ex-post contributions are desirable: the government values contributions, while the lobby is indifferent, because free entry ensures that the net return to capital will be equal to one regardless. (2) An additional reason why tariff ceilings may be superior to exact tariffs is that the presence of ex-post contributions mitigates the overinvestment problem, since positive contributions reduce net returns to capital in manufacturing.

If capital is imperfectly mobile, tariff ceilings are better than exact tariffs only if δ is relatively low. To see this, consider the extreme case in which capital is fixed. Then the only difference between tariff ceilings and exact tariffs is that the former induce ex-post contributions while the latter do not. From the ex-ante point of view, a dollar of contributions received by the government has more weight than a dollar of contributions paid by the lobby if and only if $\delta < 1$.

This proposition highlights that our model is able to explain the use of tariff ceilings, which is pervasive in real trade agreements. From another perspective, the model helps explain why trade agreements are not complete contracts, and in particular why they leave some discretion to governments. It is worth highlighting that none of the "usual" causes of contract incompleteness – e.g. nonverifiable information, costs of writing contracts, unforeseen contingencies – are present in our model. The reason why the optimal agreement may be incomplete here is that the agreement cannot specify the contributions that the lobby will have to pay in the future (as we implicitly assumed). If the agreement could specify both tariffs and contributions, a complete contract would be optimal. But since the contract cannot specify contributions, it may be optimal to leave the contract partially incomplete also in the other dimension, that is tariffs.

Given our interest in explaining the prevalence of agreements with tariff ceilings, in this section we focus on the case in which the conditions of Proposition 2 are satisfied. In a later section we will examine how results change if the agreement specifies exact tariff commitments.

Next we characterize the optimal agreement. It is instructive to start with the case of perfect capital mobility ($z = 1$), and then extend the analysis to the case of imperfect capital mobility.

2.5.1. Perfect capital mobility

Recall that the optimal agreement maximizes the ex-ante joint surplus of the two governments and the importing lobby in each sector. Given that $\widehat{x}$ is the inherited allocation of capital, this objective function can be written as:[9]

$$\Psi = aW(t,x) + xc + aW^*(t,x) + \delta[xp(t,x) - xc + (\widehat{x} - x)] \tag{2.2}$$

This expression is valid only for $x \leq \widehat{x}$, but we do not need to consider the alternative case $x > \widehat{x}$, because this can never be the case in equilibrium.

To gain better understanding about the objective function above, focus on the special case of $\delta = 1$. In this case $\Psi = J^{SR}(t,x) + aW^*(t,x) + (\widehat{x} - x)$. There are two extra terms relative to the short-run objective J^{SR}: the term $aW^*(t,x)$, which takes into account terms of trade externalities, and the term $(\widehat{x} - x)$, which captures the rents of those lobby members that will move to the N sector in the following period.

Next we derive the ex-ante objective as a reduced-form function of the tariff ceiling $\bar{t}$ and the allocation x. This is the objective function when the equilibrium of the third stage (i.e., given $\bar{t}$ and x) has been "rolled back" by backward induction. To this end, it is convenient to derive the functions $t(\bar{t},x)$ and $c(\bar{t},x)$ that give the equilibrium tariffs and contribution per unit of capital conditional on $\bar{t}$ and x, respectively.[10] To derive $t(\bar{t},x)$, notice that this is the tariff that maximizes $J^{SR}(t,x)$ subject to the constraint $t \leq \bar{t}$, and recall that $J^{SR}(t,x)$ is concave in t and maximized at $t^J(x)$. Therefore, if $\bar{t} \geq t^J(x)$ the tariff ceiling is not binding, hence $t(\bar{t},x) = t^J(x)$; and if $\bar{t} < t^J(x)$ the ceiling is binding, hence $t(\bar{t},x) = \bar{t}$. Summarizing: $t(\bar{t},x) = \min\{\bar{t}, t^J(x)\}$.

Turning to $c(\bar{t},x)$, the key observation is that, if $\bar{t} > t^W(x)$, then the home government (G) will get contributions, because its outside option in the negotiation with the lobby is given by the tariff $t^W(x)$, and the lobby has to compensate G to raise the tariff up to the ceiling $\bar{t}$; on the other hand, if $\bar{t} < t^W(x)$ then no contributions will be forthcoming, because G has no credible threat. Thus

$$c(\bar{t},x) = \begin{cases} (3a/8x)\left(t(\bar{t},x) - t^W(x)\right)^2 \text{ if } \bar{t} \geq t^W(x) \\ 0 \text{ if } \bar{t} < t^W(x) \end{cases}$$

Note that there is no loss of generality in focusing on agreements that impose binding tariff ceilings, i.e. $\bar{t} \leq t^J(x)$. Letting $\Psi(\bar{t},x)$ denote the ex-ante objective as a function of $\bar{t}$ and x, we have:

[9] Note that, since we are assuming transferable utility, we are multiplying the foreign welfare function by a for consistency.

[10] Note that we are using the same notation $c()$ as for the contribution schedule in the noncooperative equilibrium, even though this is not the same function. This is an abuse of notation, but the reader can distinguish the two functions because the first argument is t in one case and $\bar{t}$ in the other.

$$\Psi(\bar{t},x) = aW(\bar{t},x) + c(\bar{t},x)x + aW^{*}(\bar{t},x) + \delta[xp(\bar{t},x) - c(\bar{t},x)x + (\widehat{x} - x)] \tag{2.3}$$

The next step is to move back to the second stage and derive the equilibrium allocation conditional on $\bar{t}$. Clearly, if $\bar{t} > \hat{t}$ then the tariff ceiling is not binding, and the equilibrium will be given by $(\hat{t},\hat{x})$, just as if there was no agreement. On the other hand, if $\bar{t} \leq \hat{t}$ then the equilibrium allocation $x^{er}(\bar{t})$ is implicitly defined by the equal-returns condition[11]

$$p(\bar{t},x) - c(\bar{t},x) = 1$$

(see Figure 2). This establishes the following lemma:

Lemma 1. *The equilibrium allocation conditional on a tariff ceiling $\bar{t}$ is given by $x^{er}(\bar{t})$ if $\bar{t} \leq \hat{t}$, and by $\hat{x}$ otherwise.*

We now turn to the optimal trade agreement. The optimal tariff ceiling is the one that maximizes $\Psi(\bar{t},x^{er}(\bar{t}))$ for $\bar{t} \leq \hat{t}$. To write an expression for $\Psi(\bar{t},x^{er}(\bar{t}))$, recall that if $\bar{t} \geq t^{W}(x^{er}(\bar{t}))$ then there are contributions and

$$W(\bar{t},x^{er}(\bar{t})) + C = W(t^{W}(\bar{t}),x^{er}(\bar{t}))$$

On the other hand, if $\bar{t} < t^{W}(x^{er}(\bar{t}))$, then there are no contributions. Noting that $\bar{t} \geq t^{W}(x^{er}(\bar{t}))$ if and only if $\bar{t} \geq t^{W}$, then

$$\Psi(\bar{t},x^{er}(\bar{t}))|_{\bar{t}\leq\hat{t}} = \begin{cases} aW(t^{W}(\bar{t}),x^{er}(\bar{t})) + aW^{*}(\bar{t},x^{er}(\bar{t})) + \delta\widehat{x} \text{ if } \bar{t} \geq t^{W} \\ aW(\bar{t},x^{er}(\bar{t})) + aW^{*}(\bar{t},x^{er}(\bar{t})) + \delta\widehat{x} \text{ if } \bar{t} < t^{W} \end{cases} \tag{2.4}$$

where we used the fact that returns are equalized in the two sectors, which implies that the total lobby rents reduce to $\widehat{x}$.

The next result shows that $\Psi(\bar{t},x^{er}(\bar{t}))$ is maximized at free trade:

Proposition 3. *In the case of perfect capital mobility, the optimal agreement is $\bar{t}^{A} = 0$ (free trade) for all δ and a.*

Proof: We argue that free trade is better than any $\bar{t} > 0$. This is clearly true for $\bar{t} < t^{W}(\bar{t})$. Let's then focus on $\bar{t} > t^{W}(\bar{t})$. We need to show that

$$W(t^{W}(\bar{t}),x^{er}(\bar{t})) + W^{*}(\bar{t},x^{er}(\bar{t})) < W(0,x^{ft}) + W^{*}(0,x^{ft})$$

Since W^{*} is decreasing in t,

$$W(t^{W}(\bar{t}),x^{er}(\bar{t})) + W^{*}(\bar{t},x^{er}(\bar{t})) < W(t^{W}(\bar{t}),x^{er}(\bar{t})) + W^{*}(t^{W}(\bar{t}),x^{er}(\bar{t}))$$

Since $W + W^{*}$ is decreasing in t,

$$W(t^{W}(\bar{t}),x^{er}(\bar{t})) + W^{*}(t^{W}(\bar{t}),x^{er}(\bar{t})) < W(0,x^{er}(\bar{t})) + W^{*}(0,x^{er}(\bar{t}))$$

[11] Again, the notation $x^{er}(\bar{t})$ is slightly abused because this is not the same function as $x^{er}(t)$, the equal-returns condition in the absence of agreements.

But clearly

$$W(0, x^{er}(\bar{t})) + W^*(0, x^{er}(\bar{t})) < W(0, x^{ft}) + W^*(0, x^{ft})$$

which proves the claim. **QED**

We have shown that, when capital is perfectly mobile, the optimal agreement is free trade even in the presence of ex ante lobbying. Intuitively, if capital is mobile, the lobby anticipates that any rents will be dissipated by entry in the ex-post stage, and hence is not willing to pay anything to compensate the government for the long run distortions associated with protection. This will of course no longer be true when capital is imperfectly mobile.

In this model there are two motives for a trade agreement: the standard terms-of-trade (TOT) externality and the domestic commitment problem. We can disentangle the two with the following thought experiment. Consider a fictitious scenario in which the home government can commit domestically (subject to the lobby's pressures) but acts noncooperatively vis-á-vis the foreign country. More precisely, suppose that at the beginning of the game the home government and the lobby choose a tariff ceiling without cooperating with the foreign government; then capital is allocated, and then the home government and the lobby choose the tariff given the ceiling and the capital allocation.

Let $\bar{t}^{DC}$ be the tariff ceiling that would be chosen in this case. The objective is the same as in the previous case except that foreign welfare is not taken into account. So $\bar{t}^{DC}$ maximizes

$$J(\bar{t}, x^{er}(\bar{t}))|_{\bar{t} \leq \hat{t}} = \begin{cases} aW(t^W(\bar{t}), x^{er}(\bar{t})) + \delta\widehat{x} \text{ if } \bar{t} \geq t^W(\bar{t}) \\ aW(\bar{t}, x^{er}(\bar{t})) + \delta\widehat{x} \text{ if } \bar{t} < t^W(\bar{t}) \end{cases} \tag{2.5}$$

We can think of the movement from $\hat{t}$ to $\bar{t}^{DC}$ as the component of trade liberalization that is due to the domestic commitment motive, and the movement from $\bar{t}^{DC}$ to $\bar{t}^{A} = 0$ as the component due to the TOT motive. Next we characterize $\bar{t}^{DC}$ in order to say more about this decomposition.

Proposition 4. *Let (x^W, t^W) be the intersection of the $t^W(x)$ curve with the $x^{er}(t)$ curve. Then $\bar{t}^{DC} = t^W$.*

(See the proof of the more general Proposition 6 in the next section).

Note that the TOT component of the agreement, i.e. the difference $\bar{t}^{DC} - \bar{t}^{A}$, is just given by the "optimal" TOT tariff t^W (see Figure 2). This is consistent with the standard TOT models a' la Grossman-Helpman and Bagwell-Staiger, where the optimal agreement just removes TOT considerations from the countries' protection levels. It is important to note that the TOT component of the agreement is independent of politics (a).[12] On the other hand, the domestic-commitment component of the agreement, $\hat{t} - \bar{t}^{DC}$, is larger when politics are more important (a is lower). The following corollary records this result:

Corollary 1. *The terms-of-trade component of the agreement is independent of a, while the domestic-commitment component of the agreement is decreasing in a.*

We next turn to the more general case of imperfect capital mobility.

[12] Straightforward algebra reveals that $t^W = (3 - 2v)/4$.

2.5.2. Imperfect capital mobility

In this section we characterize the optimal agreement for general $z \in [0,1]$ and examine how the depth of trade liberalization varies with z.

As a first step toward characterizing the optimal agreement, it is instructive to examine the extreme case in which capital is fixed at some level x. In this case, the optimal agreement is given by

$$t^{\Psi}(x) \equiv \arg\max_{\bar{t}} \Psi(\bar{t}, x)$$

It is easy to show that the curve $t^{\Psi}(x)$ lies uniformly below the curve $t^{J}(x)$.[13] Thus, if the allocation is fixed at level x the trade agreement reduces the tariff by the amount $t^{J}(x) - t^{\Psi}(x) > 0$.

It is useful at this juncture to open a parenthesis and relate this case of fixed capital with the standard TOT story, as modeled for example in Grossman-Helpman's (GH) 1995 paper. The optimal agreement in GH's model is essentially the same as this case when $\delta = 1$. To see this, note that for $\delta = 1$, $\Psi(\bar{t}, x)$ reduces to

$$\Psi(\bar{t}, x)|_{\delta=1} = aW(\bar{t}, x) + aW^{*}(\bar{t}, x) + xp(\bar{t}, x) + (\cdot)$$

where we omit the term in $(\cdot)$ because it is constant in $\bar{t}$. This is the joint surplus of the two governments and the lobby. As in GH's model, the optimal agreement maximizes this joint surplus. Note also that in this case an exact tariff is equivalent to a tariff ceiling, because contributions wash out in the ex-ante objective.

Now consider what happens if δ is lower than one. Then the objective $\Psi(\bar{t}, x)$ becomes:

$$\Psi(\bar{t}, x) = aW(\bar{t}, x) + aW^{*}(\bar{t}, x) + \delta xp(\bar{t}, x) + (1-\delta)c(\bar{t}, x)x + (\cdot)$$

For $\bar{t}$ and x such that $\bar{t} < t^{W}(x)$ there are no contributions, and the objective is qualitatively the same as in the case $\delta = 1$ (except that the relative weight of welfare vs. profits increases); but if the tariff ceiling $\bar{t}$ exceeds $t^{W}(x)$ there are positive contributions, and the objective is qualitatively different than in the GH case. The optimal tariff ceiling may be above or below $t^{W}(x)$, depending on the parameters;[14] if it is above $t^{W}(x)$, then tariff ceilings are strictly better than exact tariffs, because contributions enter positively in the objective $\Psi(\bar{t}, x)$. Thus, if the optimal tariff ceiling is above $t^{W}(x)$, it suffices to perturb GH's model by lowering δ below one (even slightly), to find a role for tariff ceilings.

Next consider the domestic-commitment benchmark when x is fixed. The optimal tariff ceiling in this case maximizes

$$J(\bar{t}, x) = aW(\bar{t}, x) + \delta xp(\bar{t}, x) + (1-\delta)c(\bar{t}, x)x + (\cdot)$$

[13] To see this note that, for $\bar{t} \geq t^{W}$, the objective function can be written (suppressing the x argument) as $\Psi(\bar{t}) = aW(t^{W}) + aW^{*}(\bar{t}) + \delta[p(\bar{t}) - c(\bar{t})]x + (\cdot)$. Noting that the net return to capital $p(\bar{t}) - c(\bar{t})$ is maximized at $\bar{t} = t^{J}$ and that $W_{t}^{*} < 0$, it follows that the maximizer of this function is lower than t^{J}. For $\bar{t} < t^{W}$, the objective becomes $\Psi(\bar{t}) = aW(\bar{t}) + aW^{*}(\bar{t}) + \delta p(\bar{t})x + (\cdot)$. Given $\delta \leq 1$ the maximizer of this function is clearly lower than t^{J}.

[14] If the "Grossman-Helpman" tariff (i.e. the optimal agreement tariff for $\delta = 1$) is above t^{W}, then since the problem is continuous the optimum is above t^{W} also for δ sufficiently close to one.

It is not hard to show that for all δ this objective is maximized by $t^J(x)$, that is, the optimum involves no agreement at all.[15] Thus, when x is fixed, the domestic-commitment component of the agreement is nil, and the whole tariff cut is coming from the TOT component. A domestic-commitment motive for trade agreements is present only if capital is mobile ($z > 0$).

To summarize, the previous discussion allows us to disentangle the role of δ from the role of z: when the only departure from the standard GH model is that ex-ante lobbying is weaker than ex-post lobbying ($\delta < 1$), the only motive for trade agreements is still the TOT externality, *but* a role for tariff ceilings appears. On the other hand, independently of the value of δ, a domestic commitment motive for trade agreements emerges if and only if there is capital mobility (ie, $z > 0$).

We can now move to the more general case $z \in [0, 1]$. Let us start by considering the equilibrium conditional on a given tariff binding $\bar{t}$. To develop intuition, suppose that $\bar{t} < \hat{t}$ and z is small. From Lemma 1, we know that if capital were perfectly mobile, the equilibrium allocation would be the one that equalizes returns given $\bar{t}$, that is $x^{er}(\bar{t}) < \hat{x}$. But if z is small, capital will not be able to exit the import-competing sector in sufficient amount to equalize returns across sectors. The allocation will then be $x_z \equiv (1 - z)\hat{x}$ and the rate of return will be higher in the N sector. In general, the equilibrium allocation conditional on $\bar{t}$ is $\max\{x^{er}(\bar{t}), x_z\} \equiv \tilde{x}^{er}(\bar{t})$. This is simply the equal returns curve truncated at x_z. The following lemma states this result formally:

Lemma 2. *Let* $\max\{x^{er}(\bar{t}), x_z\} \equiv \tilde{x}^{er}(\bar{t})$. *The equilibrium allocation conditional on* $\bar{t}$ *is given by* $\tilde{x}^{er}(\bar{t})$ *if* $\bar{t} \leq \hat{t}$ *and by* $\hat{x}$ *otherwise.*

This lemma implies that the optimal agreement is the one that maximizes $\Psi(\bar{t}, \tilde{x}^{er}(\bar{t}))$ for $\bar{t} \leq \hat{t}$. Note that, since investors are risk neutral, what matters for the lobby is only the total expected future returns for the lobby members, which are given by $x(p-c)+(\widehat{x}-x)\cdot 1$, thus the same expression we had for $\Psi(\bar{t}, x)$ with perfect mobility is valid also with imperfect mobility. The key is that the parameter z enters the problem only through its effect on x.

We are now in a position to characterize the optimal agreement for general z. Let $t^{er}(x)$ be the inverse of $x^{er}(\bar{t})$ (in the relevant region $x^{er}(\bar{t})$ is increasing, so its inverse exists).

Proposition 5. *(i) The optimal tariff binding is given by*

$$\bar{t}^A = \begin{cases} \min(t^{er}(x_z), t^{\Psi}(x_z)) \text{ for } x_z \geq x^{ft} \\ 0 \text{ for } x_z < x^{ft} \end{cases}$$

(ii) $\bar{t}^A$ *is weakly decreasing in* z, *weakly increasing in* δ *and weakly decreasing in* a.

[15] To see this, rewrite the objective as

$$J(\bar{t}, x) = \begin{cases} aW(t^W(), x) + \delta[xp(\bar{t}, x) - xc(\bar{t}, x)] \text{ if } \bar{t} \geq t^W() \\ aW(\bar{t}, x) + \delta xp(\bar{t}, x) \text{ if } \bar{t} < t^W() \end{cases}$$

and note that (i) for $\bar{t} < t^W()$ the objective is clearly increasing in $\bar{t}$; (ii) for $\bar{t} \geq t^W()$ we have $xc(\bar{t}, x) = a[W(\bar{t}, x) - W(t^W(), x)]$; plugging this in the objective, this reduces to $xp(\bar{t}, x) + aW(\bar{t}, x)$ plus a term that doesn't depend on $\bar{t}$, which is maximized by $t^J(x)$.

Proof: The key is to show that $\Psi(\bar{t}, x^{er}(\bar{t}))$ is decreasing in $\bar{t}$. Focus first on the case $\bar{t} < t^W(\bar{t})$. Here there are no contributions, so $x^{er}(\bar{t})$ is the standard supply response to a tariff and hence the objective is the standard world welfare with mobile x, which is decreasing in $\bar{t} = 0$. What about the case $\bar{t} \geq t^W(\bar{t})$? We want to show that $F(t) \equiv W(t^W(\bar{t}), x^{er}(\bar{t})) + W^*(\bar{t}, x^{er}(\bar{t}))$ is decreasing in $\bar{t}$ for $\bar{t} > t^W(\bar{t})$. Applying the Envelope Theorem, then:

$$F'(t) = W_x(t^W(\bar{t}), x^{er}(\bar{t}))dx^{er}(\bar{t})/d\bar{t} + W_t^*(\bar{t}, x^{er}(\bar{t})) + W_x^*(\bar{t}, x^{er}(\bar{t}))dx^{er}(\bar{t})/d\bar{t}$$

Since $W_t^* < 0$ and $dx^{er}(\bar{t})/dt > 0$, it suffices to show that $W_x(t^W(\bar{t}), x^{er}(\bar{t}))$ and $W_x^*(\bar{t}, x^{er}(\bar{t}))$ are both negative. The second part is obvious given that the only effect of x on W^* is through terms of trade, and an increase in x worsens Foreign's terms of trade. As to the first part, we now show that $W_x(t^W(x), x) < 0$ for $x > x^W$. Simple derivation and some manipulation reveals that:

$$W_x(t^W(x), x) = p(t^W(x), x) - 1$$

Given the definition of x^W (i.e., $p(t^W(x^W), x^W) = 1$) then the result follows immediately.

Thus, if $t^\Psi(x_z) < t^{er}(x_z)$ then - by definition of $t^\Psi(x_z)$ - the point $(t^\Psi(x_z), x_z)$ is superior to the point $(t^{er}(x_z), x_z)$ which in turn is superior to all the other points on the curve $\widetilde{x}^{er}(\bar{t})$ for $\bar{t} > t^{er}(x_z)$. On the other hand, if $t^{er}(x_z) < t^\Psi(x)$ then all points on the vertical segment of curve $\widetilde{x}^{er}(\bar{t})$ are dominated by the point $(t^{er}(x_z), x_z)$.[16] Hence, in this case, the optimal tariff binding is simply $t^{er}(x_z)$. Of course this is true only as long as $x_z \geq x^{ft}$, otherwise the optimum is free trade.

To prove that $\bar{t}^A$ is weakly decreasing in z, note that (a) $t^{er}(x)$ is increasing in the relevant range; (b) The cross derivative $\Psi_{\bar{t}x}(\bar{t}, x)$ is positive for all $\bar{t}$ and x. This implies that $\Psi(\bar{t}, x)$ is supermodular in $\bar{t}$ and x, which in turn implies that $t^\Psi(x)$ is increasing.

To prove that $\bar{t}^A(z)$ is weakly increasing in δ and weakly decreasing in a, it is direct to check that Ψ is supermodular in $\bar{t}$ and δ, and is submodular in $\bar{t}$ and a. **Q.E.D.**

Figure 3 illustrates how the optimal agreement point A varies with z. As z increases from zero, point A travels along the $t^\Psi(x)$ schedule until it hits the equal-returns curve $t^{er}(x)$, and then travels down along the $t^{er}(x)$ curve until it reaches the free trade point. Note that both $t^\Psi(x)$ and $t^{er}(x)$ are increasing in x, so the optimal tariff binding decreases as z increases.

This result suggests an empirical prediction: trade agreements should lead to deeper trade liberalization in sectors where factors of production are more mobile. Although our basic model cannot make cross-sectoral predictions because there is a single organized sector, it would not be hard to write a multi-sector model that delivers a genuinely cross-sectoral prediction along these lines.

The last statement of Proposition 5 offers another example of how "politics" affects trade liberalization in our theory. If lobbies exert stronger ex-ante pressure on governments (δ is

[16] To see this, first note the following: if $\delta > 1/3$ then $\Psi(t, x)$ is concave in t; if $\delta < 1/3$, then $\Psi(t, x)$ is convex for $t \in [t^W(x), t^J(x)]$, but since $d\Psi(t^J(x), x)/dt < 0$, then the maximum is attained for $t \leq t^W(x)$.

The statement in the text is clearly true if $\Psi(t, x)$ is concave in t. If it is not, then as argued above the maximum is attained for $t \leq t^W(x)$, and since $\Psi(t, x)$ is concave in this interval, then it must be that $\Psi(t, x)$ is increasing for $t < t^\Psi(x)$.

higher), this will lead to trade agreements with higher maximum tariffs. In other words, just as one would expect, stronger ex-ante lobbying leads to weaker trade liberalization.

Next we want to decompose the optimal agreement into its domestic-commitment and TOT components, and examine the effect of capital mobility on each component. As in the previous section, the key step is to characterize the fictitious benchmark in which the home government (in agreement with the lobby) can commit domestically:

Proposition 6. *In the domestic-commitment benchmark,*

$$\bar{t}^{DC}(z) = \begin{cases} t^{er}(x_z) \text{ for } x_z \geq x^W \\ t^W \text{ for } x_z < x^W \end{cases}$$

Proof: Consider first the case $z = 1$. The DC agreement maximizes

$$F(\bar{t}) \equiv J(\bar{t}, x^{er}(\bar{t})) = \begin{cases} aW(t^W(\bar{t}), x^{er}(\bar{t})) + \delta\hat{x} \text{ for } \bar{t} \geq t^W \\ aW(\bar{t}, x^{er}(\bar{t})) + \delta\hat{x} \text{ for } \bar{t} < t^W \end{cases}$$

We now show that $\bar{t}^{DC} = t^W$. Consider first the case $\bar{t} < t^W$. Differentiation yields:

$$F'(\bar{t}) = aW_t(\bar{t}, x^{er}(\bar{t})) + aW_x(\bar{t}, x^{er}(\bar{t}))$$

where we have used the fact that $dx^{er}/d\bar{t} = 1$ for $\bar{t} < t^W$. Since $\bar{t} < t^W$ implies $\bar{t} < t^W(\bar{t})$, and hence $W_t(\bar{t}, x^{er}(\bar{t})) > 0$, whereas $W_t(t^W, x^{er}(t^W)) = 0$. Coming now to the second term above, differentiation yields

$$W_x(\bar{t}, x^{er}(\bar{t})) = (1/2)(1 - (v - 1) - 2\bar{t})$$

It is easy to show that this is equal to zero for $\bar{t} = t^W$, and hence is positive for $\bar{t} < t^W$. Thus, $F'(\bar{t}) > 0$ for $\bar{t} < t^W$. The previous arguments establish also that $F'(t^W) = 0$.

Now consider the case $\bar{t} > t^W$. In this case, differentiation yields (using the Envelope Theorem):

$$F'(\bar{t}) = [aW_x(t^W(\bar{t}), x^{er}(\bar{t}))]dx^{er}/d\bar{t}$$

We have already established that $dx^{er}/d\bar{t} > 0$. Hence, it is sufficient to show that $W_x(t^W(\bar{t}), x^{er}(\bar{t})) < 0$ for $\bar{t} > t^W$. This is equivalent to $W_x(t^W(x), x) < 0$ for $x > x^W$. Differentiation shows that for this case we have:

$$\begin{aligned} W_x(t^W(x), x) &= v - 1 - [3x + 1 + t^W(x)]/4 \\ &= v - 4/3 - 2x/3 \end{aligned}$$

But this is equal to zero at x^W and is clearly decreasing in x, hence the result follows immediately.

Next consider the case $z < 1$. We showed in the text that $J(\bar{t}, x)$ is maximized by $t^J(x)$. Applying the same logic as in the proof of Proposition 5 one can show that

$$\bar{t}^{DC}(z) = \begin{cases} \min(t^{er}(x_z), t^J(x_z)) \text{ for } x_z \geq x^W \\ t^W \text{ for } x_z < x^W \end{cases}$$

But $\min(t^{er}(x_z), t^J(x_z)) = t^{er}(x_z)$, hence the claim. **Q.E.D.**

Figure 4 depicts how $\bar{t}^{DC}(z)$ varies with z. The domestic-commitment component of the agreed-upon tariff cut, $\hat{t} - \bar{t}^{DC}(z)$, is clearly increasing in z. What can we say about the effect of z on the TOT component, $\bar{t}^{DC}(z) - \bar{t}^{A}(z)$? In general the answer is ambiguous, but notice that for small z the TOT component of the tariff cut *decreases* with z. To see this, consider a small increase in z from zero. Then $\bar{t}^{DC}(z)$ goes down with infinite slope, while $\bar{t}^{A}(z)$ goes down with finite slope, therefore $\bar{t}^{DC}(z) - \bar{t}^{A}(z)$ decreases. Thus we can say that the liberalization-deepening effect of factor mobility is entirely due to the domestic-commitment motive, at least for z relatively small.

3. Gradual trade liberalization

In this section we consider a continuous-time extension of the model, where the agreement can determine the path of the tariff ceiling for the future. The questions we are interested in are: Does the optimal agreement entail instantaneous liberalization, gradual liberalization, or a combination of the two? If trade liberalization has a gradual component, what does the optimal tariff path look like, and what determines the speed of liberalization?

Consider the same model as above, but now assume that time is continuous, denoted by s. As in the previous section, we assume that when the agreement opportunity arises (at time $s = 0$) the world is sitting at the long-run noncooperative equilibrium, so the capital allocation is $\hat{x}$. The trade agreement determines a (fully enforceable) future path for the tariff ceiling, $\bar{t}(s)$.

We assume that, at each point in time, a fraction z of capital-owners gets a chance to exit sector M. On the other hand, to simplify the exposition and the derivation of the main results, we assume that entry into the M sector is free: that is, owners of capital in the M sector can freely and instantaneously move to the M sector if they wish. A more symmetric specification, where there is friction in capital mobility also from the N sector to the M sector, would deliver exactly the same results, but the analysis would be more cumbersome.[17] The capital allocation x will be the physical state variable of the problem.

At each point in time after the agreement is signed, the lobby makes a take-it-or-leave-it offer (t, c) to the government, taking into account the constraints set by the agreement.

Analogously to the two-period model, we assume that the agreement maximizes the weighted ex-ante joint surplus $U^G + U^{G^*} + \delta(U^L + U^{L^*})$, where U^j is interpreted as player j's payoff in PDV terms. Here for simplicity we assume that governments and lobbies discount the future in the same way, and let $\rho > 0$ denote the common instantaneous discount rate. We note that, unlike in the two-period model, having different weights in the ex-ante joint surplus is not equivalent to having different discount rates. We chose the first of these two approaches

[17] The problem is that without free entry into the M sector, our way of modeling capital mobility would lead to a discontinuity in the law of motion of x. The rate of change dx/dt would go from $-zx$ when the value of capital in the N sector is higher than in the M sector, to $z(1-x)$ in the opposite situation. This makes the optimal control problem harder to solve, but it can be shown that the results hold also in this case.

because it makes the analysis simpler.[18]

We will focus on Markov equilibria, that is equilibria where players' strategies depend on the history only through the state variable x.[19]

The first step of the analysis is to derive the equilibrium paths of x, t and c for a *given* path of the tariff ceiling, $\bar{t}(s)$.[20] We will omit the time argument s whenever this does not cause confusion.

First note that, given the Markov restriction, the equilibrium tariff as a function of x is simply $t = \min\{\bar{t}, t^J(x)\}$. If $\bar{t} \leq t^J(x)$ the tariff ceiling is binding and hence the tariff is given by $\bar{t}$, otherwise the tariff is given by $t^J(x)$. The associated contributions are given by $c(t, x)$, just as in the static model.

To characterize the equilibrium path for x, let V^M (V^N) be the value of a unit of capital in the M (N) sector. Since there is free entry into the M sector, then in equilibrium it must be that $V^M \leq V^N$. Moreover, the following no-arbitrage conditions must hold:

$$\rho V^M = z(V^N - V^M) + \dot{V}^M + p - c \tag{3.1}$$

and

$$\rho V^N = 1 \tag{3.2}$$

To understand the first of these no-arbitrage conditions, note that the flow return to a unit of capital in the M sector (on the RHS of 3.1) is composed of three terms: the expected capital gain from moving to the N sector, $z(V^N - V^M)$, the capital gains arising from any increase in V^M, $\dot{V}^M$, plus the instantaneous profits or "dividends", $p - c$. In equilibrium this flow return must be equal to the opportunity cost of holding an asset of value V^M, given by ρV^M. The second no-arbitrage condition (3.2) is similar, except that because of free capital mobility from N to M, there cannot be capital gains to holding a unit of capital in the N sector, hence the condition is simply that the instantaneous profits of capital in N, given by 1, be equal to the opportunity cost of holding this asset, $\rho^L V^N$.

Combining the no-arbitrage equations 3.1 and 3.2 and letting $y \equiv V^M - V^N$, we obtain:

$$\dot{y} = (\rho + z)y - (p - c - 1) \tag{3.3}$$

Letting $g(t, x) \equiv p(t, x) - c(t, x) - 1$, integrating and imposing the condition $y(s) \to 0$ as

[18] We also analyzed the case in which governments and lobbies have different discount rates. It turns out that, for the problem to be concave, we need the discount rate of the governments not to be higher than the discount rate of the lobbies. Under this assumption, we find that the optimal path for the tariff ceiling has the same qualitative features as the one we characterize in this section.

[19] Our restriction to Markov equilibria rules out "reputational" equilibria, but this is quite natural since we are assuming that tariff agreements are perfectly enforceable. Reputational equilibria are "useful" for sustaining cooperative outcomes when there is no external enforcement of agreements. In our enforceable-agreement setting, it can be shown that reputational equilibria can only do "worse" than the Markov equilibrium from the point of view of the ex-ante joint surplus.

[20] Analogously to the two-period model studied in the previuos section, as long as δ is not too high then tariff ceilings are preferred to exact tariffs. We assume that this condition on δ is satisfied, so we restrict our attention to agreements that specify tariff ceilings.

$s \to \infty$,[21] we obtain:

$$y(s) = \int_s^\infty e^{-(\rho+z)(v-s)} g(t(v), x(v)) dv \leq 0 \text{ for all } s \tag{3.4}$$

It follows from the above discussion that, given a path for the maximum tariff $\bar{t}(s)$, the equilibrium conditions for $t(s), x(s)$ and $y(s)$ are the following:

(1) $t(s) = \min\{\bar{t}(s), t^J(x(s))\}$

(2) $y(s)$ satisfies 3.4 and $y(s) \leq 0$ for all s

(3) $\dot{x}(s) = -zx(s)$ if $y(s) < 0$ and $\dot{x}(s) \geq -zx(s)$ if $y(s) = 0$.

Condition $y(s) \leq 0$ in (2) is a consequence of the assumption that there is free entry into the M sector. Condition (3) simply states that if $y < 0$ then capital leaves the M sector as fast as possible, whereas if $y = 0$ then any reallocation is an equilibrium as long as it satisfies the physical restriction that capital cannot leave the M sector faster than at rate $-zx$.

We can now derive the optimal agreement $\bar{t}(s)$. The objective function is

$$\int_0^\infty e^{-\rho s} \Psi(t(s), x(s)) ds \tag{3.5}$$

where

$$\Psi(t, x) \equiv aW(t, x) + xc(t, x) + aW^*(t, x) + \delta \left[g(t, x)x + \hat{x} \right]$$

We say that a *plan* $(t(s), x(s), y(s))$ is *implementable* if there is an agreement $\bar{t}(s)$ such that $(t(s), x(s), y(s))$ is an equilibrium, i.e. satisfies conditions (1)-(3). We will look for the plan that maximizes 3.5 in the set of implementable plans, and then we will identify the agreement $\bar{t}(s)$ that implements this plan. To turn this maximization into a more standard optimal control problem, we let $u = \dot{x}$ and note that any implementable plan must satisfy equation 3.3, together with the following "relaxed" restrictions:

$$y(s) \leq 0, \ \lim_{s \to \infty} y(s) = 0, \ x(0) = \hat{x}, \text{ and } \ u(s) + zx(s) \geq 0 \ \text{ for all } \ s \geq 0 \tag{3.6}$$

Conditions 3.3 and 3.6 are *necessary* for implementability. Our approach is to maximize the objective 3.5 subject to these necessary conditions for implementability, and then verify that the solution satisfies all implementability conditions. If this is the case, then we have found the optimal plan.

To characterize the solution to this problem, we need to introduce some notation. Just as in previous sections, let $t^{er}(x)$ be the tariff that equalizes net returns to capital across sectors, which is implicitly defined by $g(t, x) = 0$, and $t^\Psi(x)$ the tariff that maximizes $\Psi(t, x)$. Also, let $x^z(s)$ represent the path of x obtained when capital exits the M sector as fast as possible until the free trade allocation is reached:

$$x^z(s) = \max\{\hat{x}e^{-zs}, x^{ft}\}$$

[21] This is a condition that there should be no "bubbles" in the asset market. We could replace this by the weaker condition that y converges to a finite value as $s \to \infty$.

In order for the problem to be well-behaved, we need to assume that a is sufficiently high. A simple and largely sufficient condition for our result to hold is $a \geq \frac{6\delta(3v-4)+1}{6(2-v)} \equiv \tilde{a}$. In the Appendix we will prove the result under a weaker (but more complicated) condition. The condition that a be sufficiently high serves essentially two purposes. First, it ensures that $t^{\Psi}(x) - t^{er}(x)$ is decreasing in x, which in turn implies that the curves $t^{\Psi}(x)$ and $t^{er}(x)$ have a unique intersection. Second, it ensures that the problem is concave, so that we can apply sufficiency conditions from optimal control theory.

We are now ready to state the main result of this section:

Proposition 7. *Assume $a \geq \tilde{a}$. The optimal agreement entails four phases:*
(i) an instantaneous drop in the tariff from $\hat{t}$ to $t^{\Psi}(\hat{x})$;
(ii) a first gradual liberalization phase in which $t(s) = t^{\Psi}(x^z(s))$, and $y(s) < 0$;
(iii) a second gradual liberalization phase in which $t(s) = t^{er}(x^z(s))$, and $y(s) = 0$;
(iv) a steady state in which the tariff is zero.
The optimal path for the allocation is given by $x(s) = x^z(s)$ for all s.

Proof: see Appendix.

This proposition states that the optimal trade agreement entails a discrete tariff cut at time zero, with the tariff dropping from $\hat{t}$ to $t^{\Psi}(\hat{x})$, which is then followed by gradual trade liberalization and exit of capital from the M sector. This gradual trade liberalization is characterized by two phases. In the first phase, the tariff is given by the best static tariff $t^{\Psi}(x^z(s))$ as a function of the evolving capital allocation, whereas in the second stage the tariff is just the one that equalizes net returns across sectors (given the capital allocation, $x^z(s)$). Note that in the first phase capitalists in the M sector want to leave as fast as possible, since the value of a unit of capital in that sector is lower than in the N sector (i.e. $y(s) < 0$); in the second phase capitalists are indifferent as to where to locate their capital (i.e. $y(s) = 0$), but the government induces exit at the fastest possible rate. After a period of adjustment, the capital stock reaches the free trade allocation, and free trade obtains thereafter.[22]

Interestingly, the immediate tariff drop that follows the signing of the trade agreement captures the terms-of-trade motive for the trade agreement, while the domestic-commitment motive is reflected in the gradual component of trade liberalization. Thus the model suggests that the gradual component of trade liberalization should be more important, relative to the instantaneous component, when the domestic-commitment motive is more important relative to the terms-of-trade motive.

Another interesting prediction of this analysis is that trade liberalization paths established in trade agreements should entail faster liberalization for sectors where exit can proceed at a

[22] We emphasize that the above result is far from being a corollary of the comparative-statics result of the static model, where we derived the optimal tariff ceiling as a function of z. As an example of the different structure of the problem, note that in this section we need a unique intersection between the t^{Ψ} curve and the t^{er} curve, a condition we did not need in the static model. In the static model, the optimal tariff ceiling always follows the lower envelope of these two curves as z increases. In the dynamic problem, if the two curves intersect more than once the optimal tariff ceilings may not follow their lower envelope. We actually find it surprising that under some simple conditions the solution of the continuous-time model mirrors exactly the comparative-statics solution of the static model.

faster pace. An interesting open question is whether this prediction is consistent with empirical observations.

4. Extensions

In this section we consider two extensions. First, what happens if governments have some bargaining power vis-á-vis their domestic lobbies? Second, how are the results affected if ex-ante lobbying is so strong (δ is so high) that Proposition 2 does not hold and exact tariffs are preferable to tariff ceilings? We address these extensions in the context of the two-period model of Section 2.

4.1. Home government's bargaining power

An assumption we have maintained thus far is that the lobby has all the bargaining power. This is a convenient assumption because in this case the government does not derive any rents from the political process, and hence - given the associated long-run losses - is always interested in a trade agreement (see Maggi and Rodríguez-Clare, 1998). If the government has some bargaining power, however, it will extract rents from the political process, and may not be interested in signing a free trade agreement. In particular, the value of a trade agreement will depend on the comparison between the rents derived by the government in the political process and the allocative distortions induced by the expectation of protection. The question arises, then, as to how our results change when we drop the extreme assumption that the lobby has all the bargaining power.

To illustrate this in the simplest way, we consider the opposite extreme, namely the case when the government has all the bargaining power. For some x consider the government and lobby negotiating a tariff above $t^W(x)$. The rents obtained by the lobby would be given by:

$$x\left[p(t,x)-p(t^W(x),x)\right]$$

Since the government has full bargaining power, it captures all these rents in the form of contributions. Thus, we have that for $t \geq t^W(x)$

$$c(t,x)=p(t,x)-p(t^W(x),x)$$

Net profits for capitalists in the Manufacturing sector are then $p(t,x)$ for $t < t^W(x)$ and $p(t^W(x),x)$ for $t \geq t^W(x)$. In other words, the $t^{er}(x)$ curve has slope one until it reaches the $t^W(x)$ curve at point W (i.e. (t^W, x^W)) and then becomes vertical. The long-run non-cooperative equilibrium is given by the intersection of the curve $p^J(x)$ and this new equal-returns (ER) curve $p^{er}(x)$, which occurs at $x = x^W$ and $p^J(x^W)$.

Consider now what happens when the governments can sign a trade agreement and there is full capital mobility ($z = 1$). We first need to examine the objective function of the two governments and the lobby for the case when the Home government has all the bargaining power vis-á-vis its own lobby. Plugging the above result for contributions into the expression for Ψ that we considered above, we obtain:

$$\begin{aligned}\Psi(\bar{t},x) &= aW(\bar{t},x)+xp(\bar{t},x)+aW^{*}(\bar{t},x)+\\ &\quad \delta[xp(t^{W}(x),x)+(\widehat{x}-x)]-xp(t^{W}(x),x)\end{aligned}$$

Just as above, let $t^{\Psi}(x)$ denote the value of $\bar{t}$ that maximizes $\Psi(\bar{t},x)$. Note that in this case $t^{\Psi}(x)$ is just the Grossman-Helpman "trade-talks" tariff, $t^{GH}(x)$, given by:

$$t^{GH}(x)=\arg\max_{t}\left(aW(t,x)+xp(t,x)+aW^{*}(t,x)\right)$$

The optimal trade agreement is simply the point that maximizes $\Psi(\bar{t},x)$ along the ER curve. There are two possibilities, depending on whether $t^{\Psi}(x)$ passes below or above the W point. Suppose first that it passes below. Then $\Psi(\bar{t},x^{W})$ increases as the the tariff ceiling falls from $t^{J}(x^{W})$ towards t^{W}. Moreover, we already have seen that $\Psi(\bar{t},x)$ increases as we move from W towards the free trade point. Thus, just as in Proposition 3, the best trade agreement entails free trade. If $t^{\Psi}(x)$ passes above the W point, on the other hand, then $\Psi(\bar{t},x^{W})$ increases as $\bar{t}$ falls from $t^{J}(x^{W})$ to $t^{GH}(x^{W})$ but then decreases as $\bar{t}$ continues to fall to t^{W}. Thus, there are two local maxima: $t^{GH}(x^{W})$ and $t=0$. Depending on parameters, either one could be the best trade agreement. In particular this depends on the height of the optimal terms-of-trade tariff t^{W}. If t^{W} is low (which is the case when trade volume is low, which in turn happens when v is relatively high), then $t^{GH}(x^{W})$ is close to $t^{J}(x^{W})$ and the W point is close to the free trade. This implies that the decline in Ψ as $\bar{t}$ falls from $t^{GH}(x^{W})$ to t^{W} is sizable and superior to the increase in welfare as we move from W to the free trade point, hence $t^{GH}(x^{W})$ is better than free trade.

To summarize the analysis so far, when $z=1$ and governments have strong bargaining power vis-á-vis their lobbies, there are two possibilities: either they slash tariffs down to the level $t^{GH}(x^{W})$ or they go all the way to free trade, with the latter being more likely when the terms-of-trade externality is stronger.

Now consider the general case of $z\in[0,1]$. As above, it is useful to consider how the optimal agreement varies as z increases from zero. To do this, it is necessary first to determine the feasible set of points (t,x) that can be attained as equilibria given some tariff binding. The equilibrium for all $\bar{t}\geq t^{W}$ entails $x=x^{W}$ whereas if $\bar{t}<t^{W}$ then the set of feasible points is given by $x^{er}(\bar{t})$ if $x^{er}(\bar{t})\geq x^{W}(1-z)$ and $x^{W}(1-z)$ otherwise. (This is a truncated curve, first vertical at $x^{W}(1-z)$ from $t=0$ until $t^{er}(x^{W}(1-z))$, then a line of slope one until the W point, and then vertical again.)

Again, there are two cases, depending on whether $t^{\Psi}(x)$ passes below or above point W. First, if $t^{GH}(x^{W})<t^{W}$ then the optimal agreement moves along $t^{GH}(x^{W}(1-z))$ as z increases from zero, until this curve crosses the ER curve. For higher z, the optimal agreement moves along $t^{er}(x^{W}(1-z))$ until z is sufficiently high that $x^{W}(1-z)=x^{ft}$. Second, if $t^{GH}(x^{W})>t^{W}$, then the optimal agreement remains at $t^{GH}(x^{W})$ for all z if this point is better than free trade. If not, then there would be a z sufficiently high at which the best agreement switches discontinuously from $t^{GH}(x^{W})$ to $t^{er}(x^{W}(1-z))$, following this curve up to free trade as z increases further.

To summarize, if the government has all the bargaining power, it is still true that trade liberalization is deeper when capital is more mobile, but it may no longer be true that with full capital mobility the trade agreement entails free trade.

4.2. Exact-tariff agreements

Here we consider the case in which δ is so high that Proposition 2 does not hold (recall that such a case can occur only if $z < 1$: if capital mobility is perfect, the result holds for any δ). As explained in Section 2.5.2, in this case restricting the analysis to agreements with tariff ceilings is no longer valid: we have to allow for the possibility that governments would prefer an agreement with exact tariffs. Of course, in this case one of our main results, namely the optimality of using tariff ceilings, no longer holds, but what about the other results? Do they still hold? This is the issue we explore in this section by characterizing the optimal strong-binding agreement.

The analysis is quite similar to that of agreements with tariff ceilings, so the presentation will be very schematic. First, note that since there are no contributions, the ex-ante joint surplus is simply:

$$\Omega(t, x) \equiv aW(t, x) + aW^*(t, x) + \delta p(t, x)x + \delta(\hat{x} - x)$$

Let $t^{\Omega}(x)$ be the level of t that maximizes $\Omega(t, x)$. It is easy to show that

$$t^{\Omega}(x) = \delta x/a$$

Next let $t_N^{er}(x)$ be the tariff that equalizes returns across sectors when there are no contributions, given x. This is given implicitly by

$$p(t, x) = 1$$

Note that both $t^{\Omega}(x)$ and $t_{NC}^{er}(x)$ are increasing functions of x. The next proposition shows that a result analogous to the one in Proposition 5 holds:

Proposition 8. *The optimal exact tariff is given by*

$$\bar{t}_S^A = \begin{cases} \min(t_N^{er}(x_z), t^{\Omega}(x_z)) \text{ for } x_z \geq x^{ft} \\ 0 \text{ for } x_z < x^{ft} \end{cases}$$

Proof: First notice that $\Omega(t_N^{er}(x), x)$ is decreasing. This follows simply from noting that:

$$\Omega(t_N^{er}(x), x) = aW(t_N^{er}(x), x) + aW^*(t_N^{er}(x), x) + \delta\hat{x}$$

which is obviously decreasing in x. This implies that the point $(t_N^{er}(x_z), x_z)$ is preferable to any point with $x > x_z$ along the ER curve with no contributions. Thus, if $t^{\Omega}(x_z) > t_N^{er}(x_z)$ then the optimal exact tariff is $t = t_N^{er}(x_z)$. But if $t^{\Omega}(x_z) < t_N^{er}(x_z)$ then by definition $\Omega(t, x)$ increases as we go down the vertical line at x_z from $t_N^{er}(x_z)$ to $t^{\Omega}(x_z)$. **Q.E.D.**

The above proposition characterizes the optimal exact tariff as a function of z. With perfect capital mobility ($z = 1$) the agreement entails free trade, just as in the case of agreements with tariff ceilings. As capital mobility (z) decreases, the optimal exact tariff increases, so the agreement entails less and less trade liberalization. To see this, note that $t^{\Omega}(x)$ and $t_N^{er}(x)$ are increasing functions of x, so $\min(t_N^{er}(x_z), t^{\Omega}(x_z))$ is a decreasing function of z.

A decomposition between the domestic commitment motive and the terms-of-trade motive can be shown to hold also in this setting just as in the case of agreements with tariff ceilings.

In summary, when δ is high all the results of the previous sections continue to hold *except* that tariff ceilings might not be optimal.

5. Conclusion

In this paper we have presented a theory that gives a prominent role to "politics" in the determination of trade agreements. This stands in contrast to the standard theory, according to which even politically-motivated governments sign trade agreements only to deal with terms-of-trade externalities. We developed a model where trade agreements may be motivated both by terms of trade externalities and by domestic commitment considerations. There are two key elements of our model: first, capital mobility and its interaction with lobbying and trade protection, which generates the problem of time inconsistency in trade policy; and second, a structural approach that distinguishes between ex-ante and ex-post lobbying, and that takes into account the effect of contributions on both the governments' payoffs and the returns to capital. The resulting framework is rich in implications. In particular, it leads to the prediction that trade agreements result in deeper trade liberalization when governments are more politically motivated, when lobbies have less influence on the negotiation of the agreement, when governments have less bargaining power vis-á-vis domestic lobbies, and when capital can move more freely across sectors. In addition, it establishes conditions under which governments prefer to sign agreements with tariff ceilings rather than exact tariff commitments, just as we observe in reality. Finally, the model predicts that trade liberalization occurs in two stages: an immediate slashing of tariffs relative to their noncooperative levels, and a subsequent gradual reduction of tariffs, where the instantaneous tariff cut is a reflection of the terms-of-trade motive for the agreement, while the domestic-commitment motive is reflected in the gradual phase of trade liberalization.

There is one assumption in our model that merits further discussion. We have assumed that international agreements are perfectly enforceable, while there are no domestic commitment mechanisms. An alternative approach would be to assume that there are no exogenous enforcement mechanisms, so that both domestic and international "agreements" have to be self-enforced through "punishment" strategies in repeated interactions. The question that arises in this case is the following: if domestic punishments are not enough to solve the domestic commitment problem, is it the case that international punishments can help governments live up to their domestic commitments?

A more precise way to formulate the above question is the following. Consider an infinitely repeated version of our model, and compare two punishment strategies: (i) a purely domestic punishment strategy where, if a country's tariff deviates from its equilibrium level, the capital allocation and the tariff in that country revert to their long-run noncooperative levels; (ii) an international punishment strategy where, if a country's tariff deviates from its equilibrium level, the capital allocation and the tariff in *both* countries revert to their long-run noncooperative levels.[23] Suppose the optimal international agreement with perfect enforcement entails tariff t^A, and the optimal domestic-commitment tariff in the absence of international agreements is $t^{DC} > t^A$. Now suppose that a purely domestic punishment strategy is not enough to sustain t^{DC}, so that the domestic commitment problem remains partially unsolved. We then ask: can an international punishment strategy sustain t^A? If this is the case, then we can say that the international reputation mechanism helps governments fully solve their domestic commitment problems. We conjecture that there will be a region of parameters for which this is the case. However, a rigorous examination of self-enforcing trade agreements will have to await further research.

[23] One could consider more severe international punishment strategies, for example a reversion to autarky. This would only strengthen the argument we are making here.

6. Appendix

Proof of Proposition 1:

We start by deriving $\hat{x}$. To do this, simply plug $t^J(x)$ into the ER curve, to get:

$$v - \frac{1}{2}(x + 1 - t^J(x)) - c(t^J(x), x) = 1$$

or

$$v - \frac{1}{2}(x + 1 - (1/3)(1 - x + 2x/a)) - x/6a = 1$$

This is an equation in x that yields a unique solution given by:

$$\hat{x} = 2a\left(\frac{3v-4}{4a-1}\right)$$

Note that the condition on a assumed in the Proposition implies that $a > 1/4$ and hence the denominator of the previous expression for $\hat{x}$ is positive, and also the numerator is positive given that we have assumed that $v > 3/2$.

The equilibrium tariff is given by

$$\hat{t} = t^J(\hat{x}) = \frac{1}{3}[1 + \frac{4-2a}{4a-1}(3v-4)]$$

Differentiating with respect to a,

$$\frac{d\hat{t}}{da} = -\frac{7}{3} \cdot \frac{3v-4}{(4a-1)^2} < 0$$

which shows the last part of the claim.

We need to show that imports are positive at the equilibrium we just found. This requires $1 - \hat{x} > \hat{t}$. Plugging in the values for $\hat{x}$ and $\hat{t}$ we find the condition

$$a > (6v-7)/6(2-v)$$

which is the condition assumed in the proposition.

For future reference we also show that the $t^{er}(x)$ curve is upward sloping. Differentiation shows

$$\frac{dt^{er}(x)}{dx} = \begin{cases} \frac{1/2+(a/12x)(3t^{er}(x)-\Delta x)-c(t^{er}(x),x)/x}{1/2-(a/4x)(3t^{er}(x)-\Delta x)} \text{ if } x > x^W \\ 1 \text{ if } x \leq x^W \end{cases} \tag{6.1}$$

$t^{er}(x)$ is clearly upward sloping if $x \leq x^W$, so let's now focus on the case $x > x^W$. It is easy to show that the denominator is positive as long as $t < t^J(x)$, while it is zero if $t = t^J(x)$ and negative otherwise. Moreover, a sufficient condition for the numerator to be positive is that

$$1/2 - c(t^J(x), x)/x > 0$$

But

$$c(t^J(x), x)/x = 1/6a$$

Hence, a sufficient condition for t^{er} to slope upwards below the t^J curve is that $a > 1/3$. But note that this implied by the assumption of the proposition. **Q.E.D.**

Proof of Proposition 2: We prove this result for the general case $z \in [0,1]$. The reader may want to read Section 5.2, where the optimal exact-tariff agreement is characterized, before reading this proof.

To prove the first claim, we show that it holds for $\bar{\delta} = 1$. We will show that the best exact tariff is dominated by a tariff ceiling set at the same level. We know that the best exact tariff agreement entails $t_S^A \leq t_S^{er}(x_z)$. We now argue that a tariff ceiling at $t = t_{sb}^A$ delivers a joint welfare that is at least as high as the joint welfare obtained under the exact tariff. To see this, note that if $t_{sb}^A \leq t^W(x_z)$ then the joint surplus in the two cases is the same, since there are no contributions. If, on the other hand, $t_{sb}^A > t^W(x_z)$, then there are contributions, and hence:

$$\begin{aligned}
\Psi(t_S^A, x_z) &= a[W(t_S^A, x_z) + W^*(t_S^A, x_z)] + (1-\delta)c(t_S^A, x_z)x_z + \delta[xp(t_S^A, x_z) + (\widehat{x} - x_z)] \\
&= \Psi^{SB}(t_S^A, x_z) + (1-\delta)c(t_S^A, x_z)x_z \\
&\geq \Psi^{SB}(t_S^A, x_z)
\end{aligned}$$

where the last inequality is ensured by $\delta \leq 1$.

To prove the second claim, note that with perfect capital mobility the net return to capital is equal to one, so the joint surplus can be written as $\Psi = a(W + W^*) + C + [\cdot]$. A similar argument as above can then be appied, noting that it is valid for any δ. **Q.E.D.**

Proof of Proposition 7:

We will prove the result under a weaker condition than the one stated in the text ($a \geq \frac{6\delta(3v-4)+1}{6(2-v)}$). Here we will assume that *either* of the following two conditions is satisfied: (i) $a \geq \frac{6\delta(3v-4)+1}{6(2-v)}$; (ii) $a \geq \max\{\frac{\delta(3v-4)}{2-v}, \frac{\delta}{2\delta-1}\}$ and $\delta > \frac{1}{2}$. Note that also this weaker restriction requires that, for given δ and v, the parameter a should be higher than some critical level.

Using $\beta(s)$ as the Kuhn-Tucker multiplier of the constraint on the control, $u(s) + zx(s) \geq 0$, and $\phi(s)$ as the multiplier function of the pure state constraint, $y(s) \leq 0$, then we have the following Hamiltonian:

$$H = e^{-\rho s}\Psi(t,x) + \beta\,[u + zx] - \phi y + \lambda_x u + \lambda_y[(\rho + z)y - g(t,x)]$$

Necessary conditions for optimality are $H_t = H_u = 0$, plus the Euler equations, $\dot{\lambda}_x = -H_x$ and $\dot{\lambda}_y = -H_y$, plus the constraints $u + zx \geq 0$, $y \leq 0$, and the complementary slackness (CS) conditions:

$$\beta \geq 0,\ \beta(u + zx) = 0, \text{ and } \phi \geq 0,\ \phi y = 0$$

$H_t = 0$ implies:

$$e^{-\rho s}\Psi_t - \lambda_y g_t = 0 \tag{6.2}$$

while $H_u = 0$ implies $\beta + \lambda_x = 0$, or

$$\beta = -\lambda_x \tag{6.3}$$

The Euler equation $\dot{\lambda}_x = -H_x$ yields:

$$\dot{\lambda}_x = -e^{-\rho s}\Psi_x - \beta z + \lambda_y g_x \tag{6.4}$$

while $\dot{\lambda}_y = -H_y$ yields:

$$\dot{\lambda}_y = \phi - \lambda_y(\rho + z) \tag{6.5}$$

Our methodology is to guess that the solution is the one stated in the Proposition and verify that it satisfies necessary and sufficient conditions for an optimum. Our conjectured solution entails three phases: the first phase goes from $s = 0$ to $s = \widetilde{s}$, and has $t = t^{\Psi}(x^z(s))$; the second phase goes from $s = \widetilde{s}$ to $s = s^{ft}$, where s^{ft} is defined by the time at which $e^{-zs}\hat{x}$ reaches x^{ft}, and entails $t = t^{er}(x^z(s))$; finally, the third stage starts at $s = s^{ft}$ and involves a steady state with $t = 0$ and $x = x^{ft}$.

To check this conjecture, we move backwards, checking first that free trade can be a steady state. From 6.2 we get:

$$\lambda_y(s) = e^{-\rho s}\delta x^{ft} \tag{6.6}$$

This implies $\dot{\lambda}_y = -\rho\lambda_y$. Plugging in 6.5 yields $\phi(s) = z\lambda_y(s)$. Hence, using 6.6 we get:

$$\phi(s) = e^{-\rho s} z\delta x^{ft} \tag{6.7}$$

Note that this clearly satisfies the condition $\phi(s) \geq 0$.

Now, since at the free trade steady state we have $u = 0$, then $u + zx = zx > 0$, and hence the CS conditions imply $\beta = 0$. Equation 6.3 then implies that

$$\lambda_x(s) = 0 \tag{6.8}$$

Plugging this (and $\dot{\lambda}_x = 0$) into the Euler equation 6.4 yields $e^{-\rho s}\Psi_x = \lambda_y g_x$, which can be shown to hold by noting that at free trade we have $\Psi_x = \delta x^{ft} g_x$ (since $W_x + W_x^* = 0$ at free trade) whereas from 6.6 we see that $\lambda_y g_x = e^{-\rho s}\delta x^{ft} g_x$.

We can now move backwards to the second phase, $s \in [\widetilde{s}, s^{ft}]$, and solve for λ_x and λ_y, and check that β, ϕ are positive (as required by the CS conditions).

Condition 6.2 can be used to solve for λ_y:

$$\lambda_y = e^{-\rho s}\Psi_t / g_t$$

Plugging this result plus $\beta = -\lambda_x$ into 6.4, and using $dt^{er}/dx = g_x/g_y$, yields:

$$\dot{\lambda}_x = \lambda_x z - e^{-\rho s}\frac{d\Psi}{dx} \mid_{g=0}$$

Now, since $e^{-\rho s}\frac{d\Psi}{dx} \mid_{g=0}$ given $x = x^z(s)$ is merely a function of time, we can denote it as $\mu(s)$, and hence we have a simple differential equation, which can be solved imposing $\lambda_x(s^{ft}) = 0$. This yields:

$$\lambda_x(s) = \int_s^{s^{ft}} \mu(v)e^{-z(v-s)}dv \tag{6.9}$$

Note that since $\frac{d\Psi}{dx} \mid_{g=0} < 0$ (this holds for $s < s^{ft}$ since $x^z(s) < x^{ft}$, see proof of Proposition 5) then $\lambda_x(s) < 0$, implying that $\beta(s) > 0$.

The only condition left to check for the second phase is that $\phi(s) \geq 0$ for $s \in [\widetilde{s}, s^{ft}]$. But from the Euler equation 6.5 we can see that this is true as long as

$$\phi(s) = \lambda_y'(s) + (\rho + z)\lambda_y(s) \geq 0$$

To verify this inequality, note that from 6.2 we get $\lambda_y = e^{-\rho s}\Psi_t/g_t$ evaluated at $t = t^{er}(s)$ (where, to simplify notation, we write $t^{er}(s)$ for $t^{er}(x^z(s))$). Letting

$$f(s) \equiv \frac{\Psi_t(t^{er}(s), x^z(s))}{g_t(t^{er}(s), x^z(s))}$$

then

$$\phi(s) = z\lambda_y(s) + e^{-\rho s}f'(s) \tag{6.10}$$

It can be shown that the assumption we made on a implies $t^{er}(x) < t^W(x)$ in the second phase of adjustment, thus there are no contributions. Differentiation shows that $f(s) = \delta x^z(s) - at^{er}(s)$. Using $dt^{er}/dx = 1$, we find:

$$f'(s) = azx^z(s)\,[1 - \delta/a]$$

Given that $\delta < a$ (this is ensured by our assumption on a), we conclude that $\phi(s) \geq 0$ for $s \in [\widetilde{s}, s^{ft}]$.

Moving now to the first phase (i.e., $s < \widetilde{s}$), our conjecture $y < 0$ implies by the CS conditions that

$$\phi(s) = 0 \tag{6.11}$$

Moreover, $t(s) = t^\Psi(x^z(s))$ implies $\Psi_t = 0$, and hence from 6.2 we get $\lambda_y = 0$ and hence $\lambda_y(\widetilde{s}) = 0$. The second Euler equation (6.5) is trivially satisfied with

$$\lambda_y(s) = 0 \tag{6.12}$$

and $\phi = 0$. To check the first Euler equation (6.4) we use $\beta = -\lambda_x$ and $\lambda_y = 0$ to obtain:

$$\dot{\lambda}_x = -e^{-\rho s}\Psi_x + \lambda_x z$$

Since Ψ_x is evaluated at $x = x^z(s)$ and $t^\Psi(x^z(s))$ then Ψ_x is just a function of time, s, hence we can write $\Psi_x(s)$. Solving the above differential equation subject to some $\lambda_x(\widetilde{s})$ yields:

$$\lambda_x(s) = \lambda_x(\widetilde{s})e^{-z(\widetilde{s}-s)} + \int_s^{\widetilde{s}} \Psi_x(v)e^{-z(v-s)-\rho v}dv \tag{6.13}$$

We must now check that $\lambda_x(s) \leq 0$, so that $\beta(s) \geq 0$. We know from 6.9 that $\lambda_x(\widetilde{s}) \leq 0$. Thus, it is sufficient to establish that $\Psi_x(s) \leq 0$ for all $s \in [0, \widetilde{s}]$. We need to do this for the case of positive contributions ($t^\Psi(x) > t^W(x)$) and the case of zero contributions ($t^\Psi(x) \leq t^W(x)$).

If there are no contributions, (using $W_x + W_x^* = (x^{ft} - x)/2$) then

$$\Psi_x = -(a/2)(x - x^{ft}) + \delta(p - 1 - x/2)$$

Since $x > x^{ft}$ then the first term is negative. To show that the second term is also negative, note that there are two cases: (1) $t^\Psi(x) \leq t^{er}(x) \leq t^W(x)$ and (2) $t^\Psi(x) \leq t^W(x) \leq t^{er}(x)$. In case (1) $p(t^{er}(x), x) = 1$, which implies that $p(t^\Psi(x), x) < 1$. In case (2) we would have $x > x^W$, which implies that $p(t^W(x), x) < 1$ and hence $p(t^\Psi(x), x) < 1$. Thus, the second term is negative.

If there are positive contributions, then

$$\Psi_x = a(W_x(t^W, x) + W_x^*(t, x)) + \delta(p - c - 1 + x(p_x - c_x))$$

Given that $t^{er}(x) > t^{\Psi}(x) > t^W(x)$, then $x > x^W$ and consequently $W_x(t^W, x) < 0$, as we showed in the proof of Proposition 5. W_x^* is always negative. $p - c - 1$ is zero at $t^{er}(x)$, hence it must be negative at $t^{\Psi}(x)$ given that $t^{\Psi}(x) < t^{er}(x)$. Hence, it suffices to show that $p_x - c_x < 0$ when evaluated at $t^{\Psi}(x)$. But $p_x - c_x = -1/2 - (1/x)\left[C_x - C/x\right]$. Since $C_x = (a/4)(t^{\Psi} - t^W) > 0$, then it is sufficient to establish that $1/2 - c(t^{\Psi}(x), x)/x > 0$. But in the proof of Proposition 1 we already established that $1/2 - c(t^J(x), x)/x > 0$. Given that $c(t^J(x), x) > c(t^{\Psi}(x), x)$, then this last inequality implies the previous one.

We have established that the conjectured solution $x = x^z(s)$ and

$$t = \begin{cases} t^{\Psi}(x^z(s)) \text{ for } s \leq \widetilde{s} \\ t^{er}(x^z(s)) \text{ for } s > \widetilde{s} \end{cases}$$

together with the implied state variable y and costate variables λ_x and λ_y given by 6.13, 6.9, 6.8 and 6.12, **??**, 6.6 in phases 1, 2, 3, respectively, and Kuhn-Tucker multipliers $\beta = -\lambda_x$ and ϕ given by 6.11, 6.10, and 6.7 in phases 1, 2, and 3, respectively, satisfy all the necessary conditions for an optimum. We now show that the conditions for sufficiency are also satisfied.

We need to show that the maximized Hamiltonian is concave in (x, y). The maximized Hamiltonian is:

$$H^0(u(x,y), t(x,y), x, y, \lambda_x, \lambda_y, s) = e^{-\rho s}\Psi(t(x,y), x) - \lambda_x z x + \lambda_y\left[(\rho^L + z)y - g(t(x,y), x)\right]$$

Clearly, it is sufficient to show that $d^2H^0/dx^2 < 0$.

Let's first analyze this in the first phase, where $t(x,y) = t^{\Psi}(x)$, $\Psi_t = 0$ and $\lambda_y = 0$. Differentiating and using $dt^{\Psi}/dx = -\Psi_{xt}/\Psi_{tt}$ yields

$$d^2H^0/dx^2 = (e^{-\rho s}/\Psi_{tt})\left(\Psi_{xx}\Psi_{tt} - \Psi_{xt}^2\right)$$

The SOC for $t^{\Psi}(x)$ requires that $\Psi_{tt} < 0$. Hence $d^2H^0/dx^2 < 0$ if and only if $\Psi_{xx}\Psi_{tt} - \Psi_{xt}^2 > 0$, which is a condition for Ψ to be concave in (x, t) at $(x^z(s), t^{\Psi}(x^z(s)))$.

We have to consider separately the cases in which there are positive and zero contributions. For the case with no contributions, we have $\Psi_{tt} = -a/2 < 0$ and

$$\Psi_{xx}\Psi_{tt} - \Psi_{xt}^2 = (1/4)\left[a^2 + 2\delta a - \delta^2\right]$$

Since $\delta < a$ then $\delta a > \delta^2$, hence the above expression is positive.

Now let's consider the case with positive contributions. This necessarily implies that $\delta > 1/2$ and $a > \delta/(2\delta - 1)$. Differentiation yields

$$\Psi_{tt} = -a/2 + 3a/4 - 3\delta a/4 = (a/2)(3/2 - 1) - 3\delta a/4 = (a/4)(1 - 3\delta)$$

which is negative given that $\delta > 1/2$, and

$$\Psi_{xx}\Psi_{tt} - \Psi_{xt}^2 = (a/48)(5a + 12\delta + \delta a)(3\delta - 1) - [\delta/2 + (1-\delta)(a/4)]^2$$

Some algebra shows that this expression is equal to

$$(a^2/12)(5\delta - 2) + (\delta/4)(2\delta - 1)(a + (a - \delta/(2\delta - 1))$$

which is positive given $\delta > 1/2$ and $a > \delta/(2\delta - 1)$ (again, this last inequality must hold when there are positive contributions, given our assumptions).

Now let's move to the second and third phases, where $t = t^{er}(x^z(s))$ and $g = 0$. We have:

$$d^2H^0/dx^2 = e^{-\rho s}\left[\Psi_{tx}dt^{er}/dx + \Psi_t d^2t^{er}/dx^2 + \Psi_{xx}\right]$$

We then need to show that the expression in the square parenthesis is negative. Since there are no contributions in the second and third phases, then $dt^{er}/dx = 1$, and hence

$$d^2H^0/dx^2 = -e^{-\rho s}\left[a/2 + \delta/2\right]$$

which is clearly negative. **Q.E.D.**

References

Bagwell, K. and R. Staiger (1999), “An Economic Theory of GATT”, *American Economic Review*, 89(1), 215-248.

Bagwell, K. and R. Staiger (2005), “Enforcement, Private Political Pressure and the GATT/WTO Escape Clause,” mimeo.

Bond, R. and J.-H. Park (2004): "Gradualism in Trade Agreements with Asymmetric Countries," *Review of Economic Studies*, 69, 379-406.

Conconi, P. and C. Perroni (2004), “Self-Enforcing International Agreements and Domestic Policy Credibility,” mimeo.

Conconi, P. and C. Perroni (2005), “The Economics of S&D Trade Regimes,” mimeo.

Furusawa, T. and E. Lai (1999): "Adjustment Costs and Gradual Trade Libealization," *Journal of International Economics*, 49, 333-61.

Grossman, G. and E. Helpman (1994), “Protection for Sale”, *American Economic Review*, 84(4), 833-850.

Grossman, G. and E. Helpman (1995a), “Trade Wars and Trade Talks”, *Journal of Political Economy*, 103(4), 675-708.

Grossman, G. and E. Helpman (1995b), “The Politics of Free-Trade Agreements”, *American Economic Review*, 85(4), 667-90.

Johnson, H. (1954), “Optimum Tariffs and Retaliation, *Review of Economic Studies*, 21(2), 142-153.

Krishna, P. (1998), “Regionalism and Multilateralism: A Political Econommy Approach,” *Quarterly Journal of Economics*, 113(1), 227-52.

Maggi, G. (1999), “The role of multilateral institutions in international trade cooperation,” *American Economic Review*, 89, 190-214.

Maggi, G. and A. Rodríguez-Clare (1998), “The Value of Trade Agreements in the Presence of Political Pressures,” *Journal of Political Economy*, vol. 106(3), 574-601.

Maggi, G. and A. Rodríguez-Clare (2005), “Import Tariffs, Export Subsidies, and the Theory of Trade Agreements,” mimeo.

Mayer, W. (1981), “Theoretical Considerations on Negotiated Tariff Adjustments,” *Oxford Economic Papers*, vol. 33, 135-153.

Mitra, D., (2002): "Endogenous Political Organization and the Value of Trade Agreements," *Journal of International Economics*, 57, 473-85.

Ornelas, E. (2004), “Rent Dissipation, Political Viability and the Strategic Adoption of Free Trade Agreements,” mimeo.

Staiger, R. (1995): "A Theory of Gradual Trade Libealization," in J. Levinsohn, A. Deardorff and R. Stern (eds.), *New Directions in Trade Theory*, University of Michigan Press, Ann Arbor, Michigan.

Staiger, R. and G. Tabellini (1987): "Discretionary Trade Policy and Excessive Protection," *American Economic Review*, 77, 823-37.

Figure 1

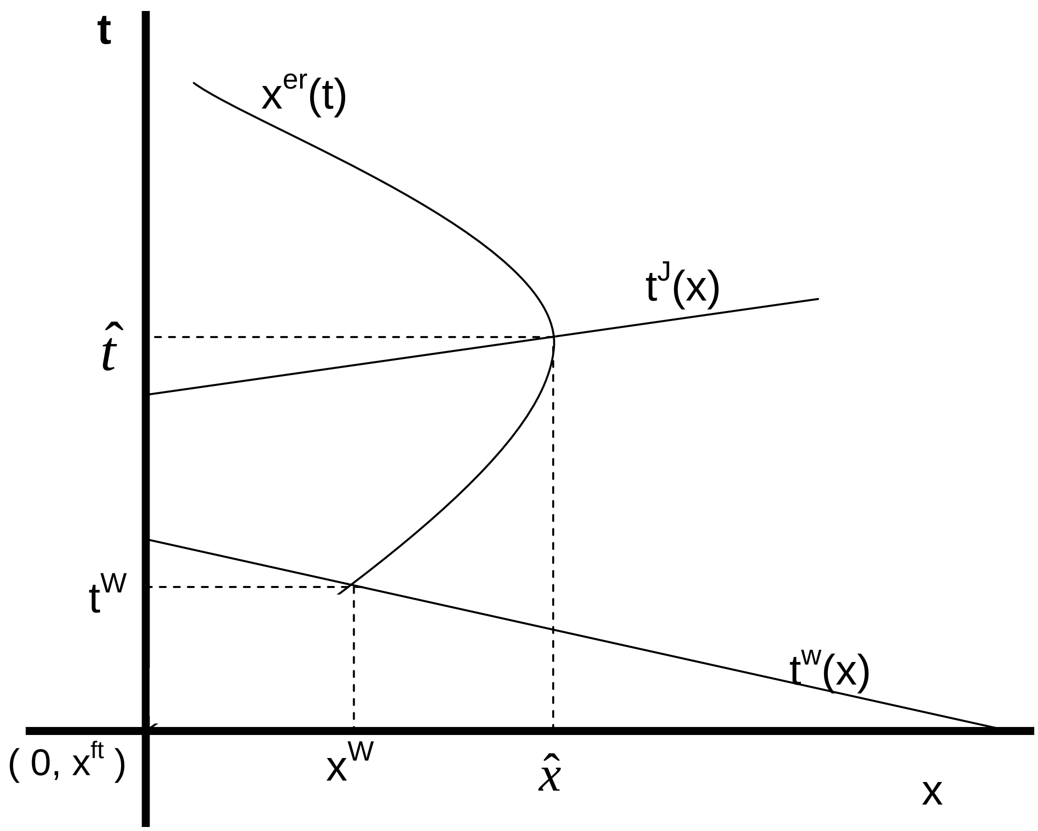

Figure 2

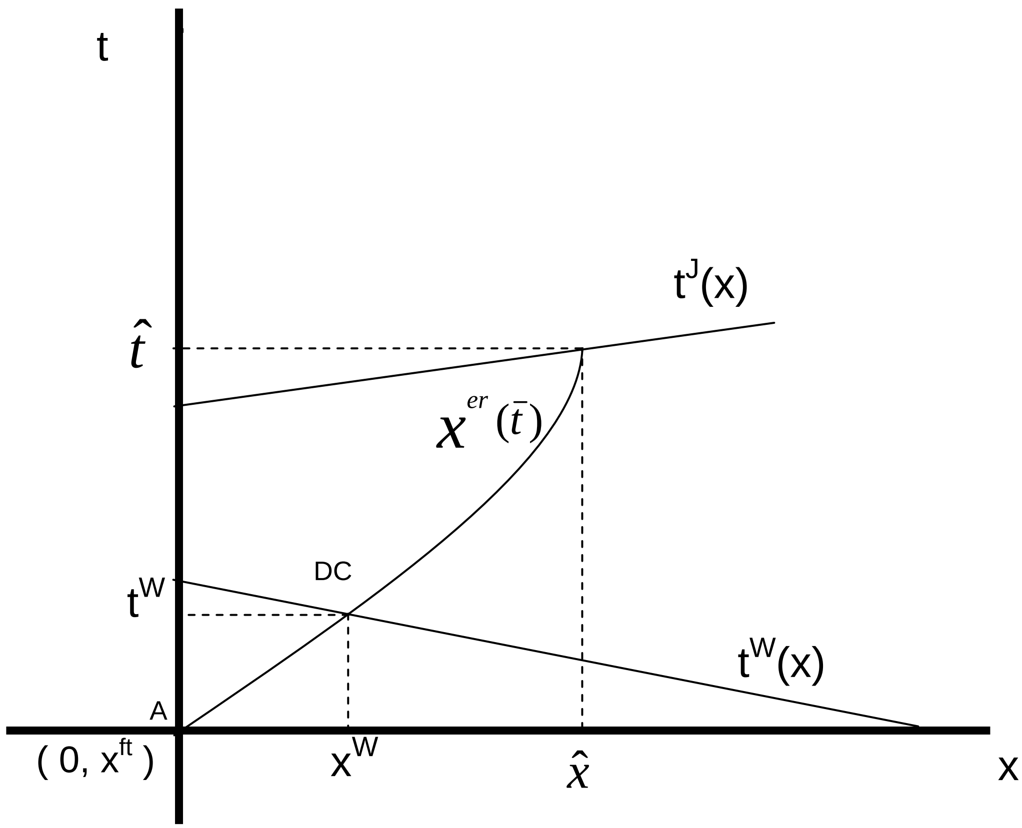

Figure 3

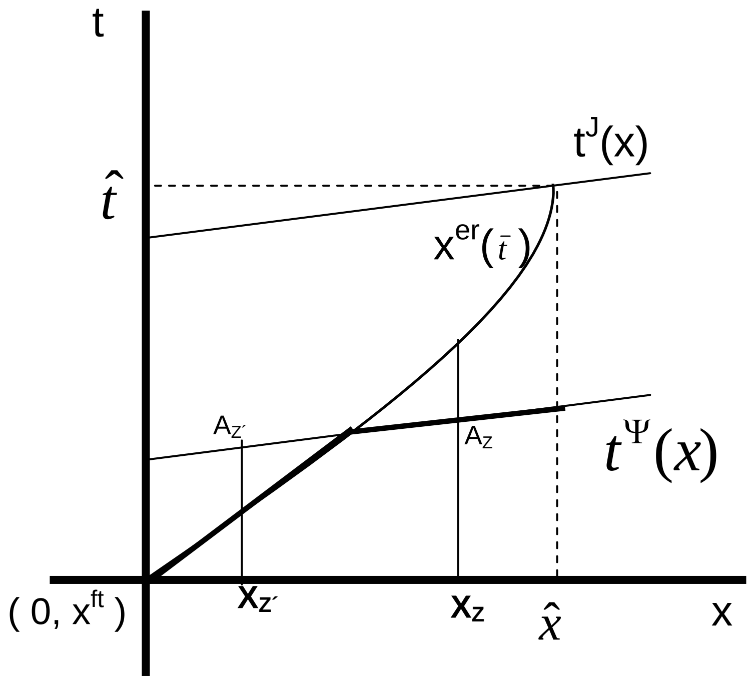

Figure 4