“Nash-in-Nash” Tariff Bargaining with and without MFN*

Kyle Bagwell† Robert W. Staiger‡ Ali Yurukoglu§

Preliminary Draft: January 2017

Abstract

We provide a theoretical framework for the analysis of the efficiency properties of bilateral tariff negotiations when transfers are unavailable and negotiations are modeled using the “Nash-in-Nash” solution concept of Horn and Wolinsky (1986). We conduct our analysis in the context of a three-country, two-good general equilibrium model of international trade, and we consider both an unrestricted negotiation setting in which discriminatory tariffs are allowed and a restricted negotiation setting in which tariffs must satisfy the MFN rule (i.e., be non-discriminatory). We allow for a general family of political-economic country welfare functions and assess efficiency relative to these welfare functions. For the unrestricted setting with discriminatory tariffs, we establish a sense in which the resulting tariffs are inefficient and too low, so that excessive liberalization occurs from the perspective of the three countries. When negotiated tariffs must satisfy the MFN rule, different cases arise. For one important case, we establish a sense in which the resulting tariffs are inefficient and too high when evaluated relative to the unrestricted set of efficient tariffs. We also compare the negotiated tariffs under the MFN rule with the MFN-constrained efficiency frontier, finding that the negotiated tariffs are generically inefficient relative to this frontier and may lead to too little or too much liberalization.

*We thank the NSF (Grant SES-1326940) for financial support.
†Department of Economics, Stanford University; and NBER.
‡Department of Economics, Dartmouth College; and NBER.
§Graduate School of Business, Stanford University and NBER

1 Introduction

Tariff negotiations between two countries can generate mutual gains by eliminating the inefficient terms-of-trade driven reductions in trade volume that occur under non-cooperative tariff setting. The effects of a bilateral trade deal, however, are not limited to the negotiating countries. Bilateral tariff cuts may also affect the welfare of other countries by altering their terms of trade. Due to this third-party externality, a tariff negotiation that is bilaterally efficient for the negotiating countries may fail to be efficient relative to the preferences of all countries.

By altering the terms of trade, bilateral tariff negotiations can affect both the export and import interests of third-party countries. To develop this point and explore the efficiency properties of bilateral tariff negotiations, we follow Bagwell and Staiger (2005, 2010, 2016) and consider a simple three-country, two-good model in which the home country exports a particular good to each of two foreign countries, where each foreign country in turn exports the other good to the home country and where the foreign countries do not trade with one another. In this setting, when the home country offers a tariff cut to one of the foreign countries, exporters in the other foreign country are disadvantaged and sell at a reduced world price. In addition, when a foreign country extends a tariff cut to the home country, the world price of the home country's export good increases, and so the other foreign country must pay a higher world price for its import good. In this model, a bilateral tariff liberalization thus induces a terms-of-trade loss for the third country, both by reducing world demand for that country's export good and by raising world demand for that country's import good.

Bagwell and Staiger (2005) show that, starting at any efficient vector of tariffs for the three countries, the home country and any one foreign country can always gain by extending bilateral tariff cuts to one another. Since the original tariffs are efficient, the bilateral tariff deal is necessarily opportunistic: the participating countries gain at the expense of the third-party foreign country, which suffers a terms-of-trade loss. This result suggests that the scope for bilaterally opportunistic trade deals is significant, and indicates that an appropriately designed multilateral trade agreement can facilitate efficient outcomes for participating countries only if some restrictions are placed on the form of bilateral tariff deals. The GATT/WTO principle of non-discrimination, as captured by the most-favored nation (MFN) rule, can be motivated in this context. The MFN rule ensures the exporters from the non-participating foreign country enjoy any tariff cut that the home country offers as part of a bilateral deal. The MFN rule, however, does not fully insulate a given foreign country from the terms-of-trade effects of a bilateral negotiation between the home country and the other foreign country, both because in the presence of the MFN rule a tariff reduction by the home country now *raises* the world price of the good exported to the home country by both foreign countries, and because a tariff reduction by the other foreign country *still* raises the world price of the good imported from the home country by both foreign countries.[1]

[1]See Bagwell and Staiger (2005). They show, however, that the MFN rule when joined with the principle of reciprocity ensures that a bilateral tariff deal does not alter the terms of trade, nor thus the welfare, of the non-participating foreign country. For further discussion of the principle of reciprocity, see Bagwell and Staiger (1999,

Bagwell and Staiger (2005) develop their findings at a general level and do not study a specific extensive-form game of tariff bargaining among the three countries. In subsequent work, Bagwell and Staiger (2010) consider rules under which efficient outcomes can be achieved in a sequential bargaining game for the three-country model when transfers are allowed, the MFN rule is required, and other restrictions on bilateral negotiations, including rules regarding reciprocity and renegotiation, may be imposed.[2] Bagwell and Staiger (2016) develop the analysis in a different direction, by characterizing the outcomes that can be achieved in a multilateral bargaining setting in which any proposed outcome must satisfy the MFN rule along with the principle of multilateral reciprocity.

In this paper, we consider the efficiency properties of tariff negotiations in the three-country model when transfers are not available, the MFN rule may or may not be imposed, and the solution of Horn and Wolinsky (1986) is utilized. We follow Bagwell and Staiger (2005, 2010, 2016) and assess efficiency relative to the preferences of countries, where countries have general political-economic welfare functions that can include both economic and distributional concerns.[3] The novel feature of our analysis here concerns the use of the Horn-Wolinsky solution. Originally developed to examine incentives for horizontal mergers in the presence of exclusive vertical relationships, the Horn-Wolinsky solution is now frequently used in the Industrial Organization literature to consider surplus division in bilateral oligopoly settings where externalities exist across firms and agreements.[4] The Horn-Wolinsky solution is sometimes referred to as a "Nash-in-Nash" solution, since it can be thought of as a Nash equilibrium between separate bilateral Nash bargaining problems. In the Horn-Wolinsky solution, any given bilateral negotiation results in the Nash bargaining solution taking as given the outcomes of the other negotiations.[5]

As emphasized in the Industrial Organization literature, an important advantage of the Horn-Wolinsky solution is that it provides a tractable foundation for quantitative analyses in a wide range of applications where negotiations are interdependent. An important limitation of the Nash-in-Nash approach, however, is that it does not require that the solution be immune to multilateral deviations. The Nash-in-Nash approach is most directly interpreted in terms of a "delegated agent" model where a player (e.g., a firm in a merger analysis, or a country in a tariff negotiation) may be involved in multiple bilateral negotiations while relying on separate agents for each negotiation, where agents are unable to communicate with one another during the negotiation process. This interpretation may be strained in many settings of interest, including tariff negotiations, where

2002, 2005, 2016).

[2] See also Chan (2015).

[3] We thus include leading political-economy models of trade policy as well as the possibility that countries maximize national income. See Bagwell and Staiger (1999, 2002) for further discussion. For simplicity, in this paper, we refer to "government welfare" as "country welfare."

[4] For example, Crawford and Yurukoglu (2012) and Crawford, Lee, Whinston, and Yurukoglu (2016) explore negotiations between cable television distributors and content creators, while Grennan (2013), Gowrisankaran, Nevo, and Town (2015), and Ho and Lee (2016) consider negotiations between hospitals, medical device manufacturers, and health insurers.

[5] The Nash-in-Nash approach is broadly related to the pairwise-proof requirements that are directly imposed in contracting equilibria (Cremer and Riordan, 1987) or indirectly implied under the requirement of "passive" beliefs in vertical contracting models (McAfee and Schwartz, 1994 and Hart and Tirole, 1990). See McAfee and Schwartz (1995) for further discussion.

within-negotiation communication between agents associated with the same player would seem feasible.[6,7] Despite this limitation, on balance the tractability advantages of the Horn-Wolinsky solution make it a potentially valuable tool, albeit only one such tool, for exploring the efficiency properties of bilateral tariff negotiations in various institutional environments. The purpose of our paper is to provide a theoretical foundation for such explorations.

In the context of the three-country tariff negotiation considered here, the Nash-in-Nash approach is captured with a representation in which the home country simultaneously negotiates with each foreign country, where the bargaining outcome in each bilateral negotiation is determined by the Nash bargaining solution and under the assumption that the Nash bargaining outcome will be successfully achieved in the other bilateral negotiation. If we were to interpret this approach in terms of a delegated agent model, then we might imagine that the home country sends one agent to negotiate with one foreign country and another agent to negotiate with the other foreign country, where the home-country agents each maximize a common home-country welfare function welfare but are unable to communicate with each other during the course of their respective bilateral negotiations. Within this general framework, we can then consider both an unrestricted negotiation environment, in which the home country can select discriminatory tariffs across its foreign trading partners, and an MFN-constrained environment, in which the home country is restricted to apply the same import tariff to exports from both foreign countries. In the latter case, and in line with GATT/WTO practice, we assume that the home country negotiates with only one foreign country, referred to as the "principal supplier," and then extends any tariff change to the other foreign country on an MFN basis.

We begin our analysis by establishing conditions for the existence of an interior Horn-Wolinsky solution when discriminatory tariffs are allowed (i.e., in the unrestricted environment). We then characterize the efficiency properties of an interior Horn-Wolinsky solution. For the setting in which discriminatory tariffs are allowed, we establish a sense in which the resulting tariffs are inefficient and too low, so that excessive liberalization occurs from the perspective of the three countries. Formally, we start at an interior Horn-Wolinsky solution and explicitly construct a perturbation under which all four tariffs (two tariffs for the home country, and one tariff for each foreign country) are raised in a manner that generates welfare gains for each of the three countries.

Having thus constructed a particular tariff-increasing perturbation that is sufficient for Pareto gains for all countries, we then consider the necessary features of any Pareto-improving tariff perturbation, where we again start with the interior Horn-Wolinsky solution for the setting with

[6] Collard-Wexler, Gowrisankaran and Lee (2016) develop micro-foundations for the Nash-in-Nash approach for negotiations that concern bilateral surplus division. The trade application that we consider here is different, however, in that negotiations are over tariffs (rather than lump-sum transfers) which have direct efficiency consequences.

[7] In their study of the GATT Torquay Round, Bagwell, Staiger and Yurukoglu (2016) highlight the impact of a failed US/UK negotiation on other bilateral negotiations within the round. The Nash-in-Nash approach would not seem well-suited for a study of this behavior, for example. More generally, the Nash-in-Nash approach does not seem well-suited for a multilateral bargaining setting in which any proposed outcome must satisfy the MFN rule and the principle of multilateral reciprocity. As Bagwell and Staiger (2016) discuss, when these requirements are strictly imposed, a home-country proposal for greater liberalization in one bilateral relationship is feasible only if the proposal calls for less liberalization in the other bilateral relationship.

discriminatory tariffs. Given that the model allows for four tariffs, and that each country has a direct interest in each of the four tariffs, we would not expect to find that Pareto gains are possible only if each individual tariff is perturbed toward a higher value. Indeed, a country experiences an externality from a bilateral negotiation to which it is not party if and only if its terms of trade are altered as a consequence of the tariff changes in that negotiation. Building on this perspective, we show that, if all countries enjoy weak welfare gains under a perturbation from an interior Horn-Wolinsky solution, then the perturbation cannot be characterized by "opportunistic" bilateral tariff changes in both bilateral relationships, where opportunistic bilateral tariff changes are bilateral tariff changes that impose a welfare (i.e., terms-of-trade) loss on the non-participating country. Using this finding, we further show that, if at least one country strictly gains under such a perturbation, then there must exist a bilateral relationship in which at least one tariff rises while the other tariff in that relationship can fall but only to a limited extent.

We next consider the implications of the MFN rule for efficiency. As mentioned above, when the MFN rule is imposed on the Nash bargaining solution, we assume that the home country negotiates only with its principal supplier. We refer to the resulting solution as the Horn-Wolinsky MFN solution. We first examine the efficiency properties of this solution when efficiency is evaluated relative to the full space of (potentially discriminatory) tariffs. We show that different cases arise, depending on whether or not the home country would prefer more trade with its principal supplier country when taking as given the world price determined under the MFN Horn-Wolinsky solution. If the home country prefers greater trade in this sense, we can construct a perturbation from the Horn-Wolinsky MFN solution under which all countries gain and all four tariffs are reduced. For this case at least, we thus establish a sense in which the Horn-Wolinsky MFN tariffs are inefficient in the sense of being too high. More generally, our findings indicate that the MFN rule provides a partial counterbalance to the forces that result in inefficiently low tariffs when discriminatory tariffs are permitted.[8]

Finally, we consider the efficiency properties of the Horn-Wolinsky MFN solution when efficiency is evaluated relative to the MFN-constrained efficiency frontier. Drawing on Bagwell and Staiger's (2005) characterization of the MFN-constrained efficiency frontier, we show that the Horn-Wolinsky MFN solution is generically inefficient relative to the MFN-constrained efficiency frontier and may lead to either too little liberalization or too much liberalization relative to MFN-constrained efficient levels.

Overall, our analysis provides a set of efficiency characterizations for bilateral tariff negotiations while using the Horn-Wolinsky solution. The associated Nash-in-Nash bargaining model is a workhorse model in applied work in Industrial Organization that studies surplus division in bilateral oligopoly, and our work here provides a theoretical foundation for related applications in the context of bilateral tariff negotiations. Relative to work in Industrial Organization, a novel feature of our analysis is that we study bilateral relationships with two-way interactions (each country

[8] As we discuss in Section 6, we are unable to provide necessary features of Pareto-improving perturbations for the MFN setting, since the Horn-Wolinsky MFN tariffs leave unrestricted the tariff of the non principal supplier foreign country.

both sells to and buys from its trading partner). Another important feature of our work is that we characterize the efficiency properties of the Horn-Wolinsky solution under different bargaining rules. Our focus on efficiency is appropriate given our aim to study the welfare implications of bilateral tariff negotiations. We emphasize again, though, that the Horn-Wolinsky solution also has limitations, since the associated delegated agent interpretation is restrictive and may be more naturally applied in some settings than others.

The paper is organized as follows. Section 2 presents the basic three-country model of trade that we analyze. As we discuss there, we consider a general family of welfare functions for countries. Section 3 contains our results concerning the existence of an interior Horn-Wolinsky solution. For the setting with discriminatory tariffs, Section 4 contains our construction of a Pareto-improving perturbation relative to an interior Horn-Wolinsky solution. Section 5 provides related findings concerning the necessary features of Pareto-improving perturbations for this setting. Section 6 contains our constructions of Pareto-improving perturbations relative to the Horn-Wolinsky MFN solution, while Section 7 considers the relationship of the Horn-Wolinsky MFN solution to the MFN-constrained efficiency frontier. Section 8 contains a brief discussion of extensions related to MFN bargaining beyond the principal supplier rule and also to the possibility of renegotiation. Section 9 concludes.

2 Trade Model

We employ the same three-country model of trade as studied by Bagwell and Staiger (2005).[9] In this section, we briefly summarize this model and highlight some of the features that we build upon in later sections.

The model features one home country and two foreign countries, which trade two goods, x and y, that are normal goods in consumption and produced in perfectly competitive markets under conditions of increasing opportunity costs. Each foreign country trades only with the home country, which imports good x from each of the two foreign countries in exchange for exports of good y. An implication of this trading structure is that the home country is the only country that has the opportunity to set discriminatory tariffs. As usual, foreign country variables are denoted with an asterisk.

The home local relative price is denoted as $p \equiv p_x/p_y$, where p_x (p_y) is the local price in the home country of good x (y). The local relative price in foreign country $*i$, $i = 1, 2$, is likewise denoted as $p^{*i} \equiv p_x^{*i}/p_y^{*i}$. The ad valorem import tariff that the home country applies to exports of good x from foreign country $*i$ is denoted as t^i, and the ad valorem import tariff that foreign country $*i$ applies to exports of good y from the home country is denoted as t^{*i}. Throughout, we assume that tariffs are non-prohibitive. The world relative price for trade between the home country and foreign country $*i$ is denoted as $p^{wi} \equiv p_x^{*i}/p_y$. The world and local prices are related as $p = \tau^i p^{wi} \equiv p(\tau^i, p^{wi})$ and $p^{*i} = p^{wi}/\tau^{*i} \equiv p^{*i}(\tau^{*i}, p^{wi})$, where $\tau^i = 1 + t^i$ and $\tau^{*i} = 1 + t^{*i}$. The

[9] For further development of the model, see Bagwell and Staiger (1999, 2002, 2010, 2016).

implied linkage relationship is then that $p^{wi} = [\tau^i/\tau^j]p^{wj}$. Under MFN tariffs, $\tau^1 = \tau^2$ and hence a single world price emerges: $p^{wi} \equiv p^w$ for $i = 1, 2$. By contrast, under discriminatory tariffs, $\tau^1 \neq \tau^2$ and so $p^{wi} \neq p^{wj}$.

We are now prepared to define the terms of trade for each country. In this model, foreign country $*i$'s terms of trade are given by p^{wi}. The home country's bilateral terms of trade with foreign country $*i$ are likewise defined as $1/p^{wi}$. The home country's multilateral terms of trade can then be defined using a trade-weighted average of its bilateral terms of trade. Formally, we define the home country's multilateral terms of trade as $1/T$, where

$$T(p^{*1}, p^{*2}, p^{w1}, p^{w2}) \equiv \sum_{i=1,2} s^{*i}(p^{*1}, p^{*2}, p^{w1}, p^{w2}) \cdot p^{wi}$$

with

$$s^{*i}(p^{*1}, p^{*2}, p^{w1}, p^{w2}) \equiv E^{*i}(p^{*i}, p^{wi}) / \sum_{j=1,2} E^{*j}(p^{*j}, p^{wj})$$

and where $E^{*i}(p^{*i}, p^{wi})$ denotes exports of good x from foreign country $*i$ to the home country. We assume that the share functions, $s^{*i}(p^{*1}, p^{*2}, p^{w1}, p^{w2})$, are continuously differentiable.

We observe that, under MFN tariffs, $T = p^{wi} \equiv p^w$; thus, the home country's bilateral and multilateral terms of trade are equal under MFN tariffs. Intuitively, under MFN tariffs, the home country's bilateral terms of trade are invariant across foreign trading partners, and so the home country's multilateral terms of trade takes this common value as well. A discriminatory tariff policy, on the other hand, implies that, for all i, $T \neq p^{wi}$. To see why, suppose that the home country imposes a higher tariff on exports from foreign country $*i$, so that $\tau^i > \tau^j$ for $j \neq i$. Since $p = \tau^i p^{wi}$ for $i = 1, 2$, it then follows that $p^{wi} < p^{wj}$, and so the home country enjoys a better bilateral terms of trade with foreign country $*i$. The home country's multilateral terms of trade is then $1/T$, where T is a trade-weighted average of the home country's bilateral terms of trade.

As discussed further in Bagwell and Staiger (2005), trade-balance and market-clearing conditions may be stated using this notation. Let the market-clearing world prices be denoted as $\widetilde{p}^{wi}(\boldsymbol{\tau})$, for $i = 1, 2$, where $\boldsymbol{\tau} \equiv (\tau^1, \tau^2, \tau^{*1}, \tau^{*2})$. We assume that $\widetilde{p}^{wi}$ is a continuously differentiable function. For the general case in which home may select discriminatory tariffs, we assume further that $\widetilde{p}^{wi}$ is increasing in τ^j, τ^{*j} and τ^{*i}, and is decreasing in τ^i. Thus, foreign country $*i$ suffers a terms of trade loss when τ^j and τ^{*j} drop via bilateral negotiations between the home country and foreign country $*j$, for $j \neq i$. For the case in which the home country selects MFN tariffs, with $\tau \equiv \tau^1 = \tau^2$, we write the market-clearing world price as $\widetilde{p}^w(\tau, \tau^{*1}, \tau^{*2})$. We assume for this case that $\widetilde{p}^w$ is increasing in τ^{*1} and τ^{*2} and is decreasing in τ, which is to say that home enjoys a terms of trade gain when it raises τ or when τ^{*i} drops for some $i = 1, 2$.

With the underlying model of trade defined in this way, we assume that the home-country and foreign-country $*i$ preferences, respectively, are continuously differentiable functions that can be written as $w(p, T)$ and $w^{*i}(p^{*i}, \widetilde{p}^{wi})$, with all prices evaluated at market-clearing levels and thus detemined by the underlying vector of tariffs, $\boldsymbol{\tau}$.[10] The key assumption that we impose on preferences is that each country benefits from a terms of trade gain: $\partial w/\partial T < 0$ and $\partial w^{*i}/\partial \widetilde{p}^{wi} > 0$. As Bagwell and Staiger (1999, 2002) discuss in greater detail, this assumption holds when countries maximize national income, and it is also satisfied by the leading political-economy models of trade policy.[11]

Notice that, for a given country, welfare is affected by changes in the tariffs of other countries only if those changes lead to a change in the given country's terms of trade, where for the home country the relevant terms-of-trade measure is its multilateral terms of trade. An externally generated terms-of-trade change affects a country's welfare directly and also indirectly, through the induced changes in local prices. For example, a change in τ^i affects w^{*j} only if it affects $\widetilde{p}^{wj}$, where any change in $\widetilde{p}^{wj}$ directly impacts w^{*j} and also indirectly impacts w^{*j} through the induced change in $p^{*j} = \widetilde{p}^{wj}/\tau^{*j}$. Following Bagwell and Staiger (2005), we assume just above that the direct effect of a terms-of-trade change on a country's welfare can be signed: holding fixed a country's local price, a country's welfare rises when the country experiences an improvement in its terms of trade. We make no assumptions, however, about the impact of a local-price change on a country's welfare.

Bagwell and Staiger (1999, 2002, 2005) define the *politically optimal tariffs* as the vector of tariffs that satisfies $\partial w/\partial p = 0 = \partial w^{*i}/\partial p^{*i}$ for $i = 1, 2$. With discriminatory tariffs allowed, this definition imposes three conditions on four tariffs, so that there are many politically optimal tariffs. The *politically optimal MFN tariffs* are the politically optimal tariffs for which $\tau^1 = \tau^2$. These tariffs are uniquely defined. Bagwell and Staiger (1999, 2002) show that politically optimal tariffs are efficient relative to the country welfare functions w, w^{*1} and w^{*2} if and only if the tariffs also satisfy the MFN rule.

We can now represent a country's welfare in reduced form as $W(\boldsymbol{\tau}) \equiv w(p, T)$ and $W^{*i}(\boldsymbol{\tau}) \equiv w^{*i}(p^{*i}, \widetilde{p}^{wi})$, where it follows under our assumptions that $W(\boldsymbol{\tau})$ and $W^{*i}(\boldsymbol{\tau})$ are continuously differentiable functions. In view of the local- and world-price effects of changes in tariffs, Bagwell and Staiger (2005) do not impose general restrictions on the relationships between tariffs and reduced-form country welfare functions. They do impose some additional structure, however, on these relationships when tariffs are efficient, where efficiency is evaluated relative to country welfare functions. Specifically, for efficient tariffs, Bagwell and Staiger (2005) assume that

$$\begin{aligned} \frac{\partial W}{\partial \tau^i} &> 0 \text{ and } \frac{\partial W^{*i}}{\partial \tau^{*i}} > 0 \\ \frac{\partial W}{\partial \tau^{*i}} &< 0 \text{ and } \frac{\partial W^{*i}}{\partial \tau^i} < 0 \\ \frac{\partial W^{*i}}{\partial \tau^{*j}} &> 0 \text{ and } \frac{\partial W^{*i}}{\partial \tau^j} > 0. \end{aligned} \tag{1}$$

[10] With the market-clearing world prices determined as $\widetilde{p}^{wi}(\boldsymbol{\tau})$, the market-clearing local prices are given as $p(\tau^i, \widetilde{p}^{wi}(\boldsymbol{\tau}))$ and $p^{*i}(\tau^{*i}, \widetilde{p}^{wi}(\boldsymbol{\tau}))$ and are thus also determined by tariffs. Using the definition of T provided in the text, it is now straightforward to see that the market-clearing value for T can also be expressed as a function of tariffs.

[11] Political-economy models allow for distributional concerns. For simplicity, we refer to "government welfare" as "country welfare" in this paper.

Under this assumption, Bagwell and Staiger (2005) show that, at any efficient tariff vector,

$$-\frac{\partial W}{\partial \tau^{*i}}/\frac{\partial W}{\partial \tau^{i}} > -\frac{\partial W^{*i}}{\partial \tau^{*i}}/\frac{\partial W^{*i}}{\partial \tau^{i}} > 0 > -\frac{\partial W^{*j}}{\partial \tau^{*i}}/\frac{\partial W^{*j}}{\partial \tau^{i}}. \quad (2)$$

This means that, at any efficient tariff vector, the home country and foreign country $*i$ could lower τ^i and τ^{*i} in such a fashion as to enjoy mutual gains while imposing a terms-of-trade loss and indeed a welfare loss on foreign country $*j$. In effect, starting at any efficient tariff vector, the home country and foreign country $*i$ can move into a downward lens of mutual gain while generating a welfare loss for foreign country $*j$. In this sense, when discriminatory tariffs are allowed, any efficient point is vulnerable to bilateral opportunism.

The material presented in the section represents the modeling framework on which our analysis in subsequent sections builds. Each of the sections below, however, is self-contained as regards the additional structure that is placed on the manner in which tariffs affect reduced-form country welfare functions; in particular, in our analysis below, we do not maintain the assumption (1) and the associated characterization in (2) of efficient tariffs. Instead, we will impose additional structure on reduced-form country welfare functions explicitly and as needed in the analysis that follows. Bagwell and Staiger (2005) also consider MFN-efficient tariffs, which are the efficient tariffs under the restriction of the MFN rule: $\tau \equiv \tau^1 = \tau^2$. We postpone further discussion of this scenario and the relevant background findings until Section 7.

3 Horn-Wolinsky Solution: Existence Results

In this section, we define the Horn-Wolinsky solution for our trade application with simultaneous bilateral bargaining. We focus here on the case where the home country is free to use discriminatory tariffs. We also provide results concerning the existence of a Horn-Wolinsky solution in our application.

Our existence results build on standard fixed point theorems. In particular, following the application of Kakutani's fixed point theorem by Nash (1950) and Debreu (1952), we know that a pure strategy Nash equilibrium exists for a normal form game if the strategy sets are non-empty, convex and compact subsets of Euclidian space and each player's payoff function is continuous in the actions of all players and quasi-concave in that player's own action.[12] Our goal is to apply this theorem to our Horn-Wolinsky setting.

To this end, we first define the Horn-Wolinsky solution concept for our tariff bargaining application. Let us fix an initial tariff vector, $\boldsymbol{\tau}_0 \equiv (\tau_0^1, \tau_0^2, \tau_0^{*1}, \tau_0^{*2})$, which we take to be exogenous for our present purposes. Consider the bilateral negotiation between the home country and foreign country $*1$. Beginning from their initial tariffs τ_0^1 and τ_0^{*1} and taking τ^2 and τ^{*2} as given, the home

[12] For statements, see Friedman (1990, Chapter 3), Vives (1999, page 16), and Dasgupta and Maskin (2015).

country and foreign country $*1$ choose their Nash bargaining tariffs to solve

$$\max_{(\tau^1,\tau^{*1})\in S} \Delta W^1(\tau^1,\tau^{*1},\tau^2,\tau^{*2};\tau_0^1,\tau_0^{*1})\cdot \Delta W^{*1}(\tau^1,\tau^{*1},\tau^2,\tau^{*2};\tau_0^1,\tau_0^{*1}), \tag{3}$$

where $S \equiv [\underline{\tau},\overline{\tau}]^2$ with $(\underline{\tau},\overline{\tau}) \in \Re^2$ with $0<\underline{\tau}<\overline{\tau}$,

$$\Delta W^1(\tau^1,\tau^{*1},\tau^2,\tau^{*2};\tau_0^1,\tau_0^{*1}) \equiv [W(\tau^1,\tau^{*1},\tau^2,\tau^{*2}) - W(\tau_0^1,\tau_0^{*1},\tau^2,\tau^{*2})]^{\alpha}$$

and

$$\Delta W^{*1}(\tau^1,\tau^{*1},\tau^2,\tau^{*2};\tau_0^1,\tau_0^{*1}) \equiv [W^{*1}(\tau^1,\tau^{*1},\tau^2,\tau^{*2}) - W^{*1}(\tau_0^1,\tau_0^{*1},\tau^2,\tau^{*2})]^{1-\alpha}.$$

Similarly, beginning from their initial tariffs τ_0^2 and τ_0^{*2} and taking τ^1 and τ^{*1} as given, home and foreign country $*2$ choose their Nash bargaining tariffs to solve

$$\max_{(\tau^2,\tau^{*2})\in S} \Delta W^2(\tau^1,\tau^{*1},\tau^2,\tau^{*2};\tau_0^2,\tau_0^{*2})\cdot \Delta W^{*2}(\tau^1,\tau^{*1},\tau^2,\tau^{*2};\tau_0^2,\tau_0^{*2}), \tag{4}$$

where

$$\Delta W^2(\tau^1,\tau^{*1},\tau^2,\tau^{*2};\tau_0^2,\tau_0^{*2}) \equiv [W(\tau^1,\tau^{*1},\tau^2,\tau^{*2}) - W(\tau^1,\tau^{*1},\tau_0^2,\tau_0^{*2})]^{\alpha}$$

and

$$\Delta W^{*2}(\tau^1,\tau^{*1},\tau^2,\tau^{*2};\tau_0^2,\tau_0^{*2}) \equiv [W^{*2}(\tau^1,\tau^{*1},\tau^2,\tau^{*2}) - W^{*2}(\tau^1,\tau^{*1},\tau_0^2,\tau_0^{*2})]^{1-\alpha}.$$

Given $S \equiv [\underline{\tau},\overline{\tau}]^2$ with $(\underline{\tau},\overline{\tau}) \in \Re^2$ and $0<\underline{\tau}<\overline{\tau}$, and for $(\tau_0^1,\tau_0^{*1},\tau_0^2,\tau_0^{*2}) \in S^2$, we now say that a tariff vector $\tau_{hw} \equiv (\tau_{hw}^1,\tau_{hw}^{*1},\tau_{hw}^2,\tau_{hw}^{*2}) \in S^2$ is a *Horn-Wolinsky solution* if $(\tau_{hw}^1,\tau_{hw}^{*1})$ solves (3) given $(\tau^2,\tau^{*2}) = (\tau_{hw}^2,\tau_{hw}^{*2})$ and $(\tau_{hw}^2,\tau_{hw}^{*2})$ solves (4) given $(\tau^1,\tau^{*1}) = (\tau_{hw}^1,\tau_{hw}^{*1})$. In other words, $\tau_{hw} \equiv (\tau_{hw}^1,\tau_{hw}^{*1},\tau_{hw}^2,\tau_{hw}^{*2}) \in S^2$ is a Horn-Wolinsky solution if it simultaneously solves the programs given in (3) and (4). The Horn-Wolinsky solution can thus be interpreted as a "Nash-in-Nash" solution, since each bilateral pair selects its Nash bargaining solution under the assumption that the other bargaining pair does as well.

To establish conditions for the existence of a solution to the Horn-Wolinsky model, let us now define a normal form game with two players. Each player $i=1,2$ has a strategy $s^i \equiv (\tau^i,\tau^{*i})$, where $s^i \in S \equiv [\underline{\tau},\overline{\tau}]^2$ with $(\underline{\tau},\overline{\tau}) \in \Re^2$ and $0<\underline{\tau}<\overline{\tau}$. Player i has the payoff function $g^i(s^1,s^2)$, where $i \neq j$ and

$$\begin{aligned} g^1(s^1,s^2) &\equiv \Delta W^1(s^1,s^2;s_0^1)\cdot \Delta W^{*1}(s^1,s^2;s_0^1) \\ g^2(s^1,s^2) &\equiv \Delta W^2(s^1,s^2;s_0^2)\cdot \Delta W^{*2}(s^1,s^2;s_0^2), \end{aligned}$$

where $s_0^i \equiv (\tau_0^i,\tau_0^{*i}) \in S$, $i=1,2$. The strategy sets are non-empty, convex and compact subsets of Euclidian space. The payoff functions are also continuously differentiable and thus continuous in tariffs; hence, our maintained assumptions ensure that the payoff functions are continuous in s^1 and s^2 for $s^i \in S$, $i=1,2$. Hence, we have the existence of a pure strategy Nash equilibrium for the Horn-Wolinsky model if, for $i=1,2$, $g^i(s^1,s^2)$ is quasiconcave in $s^i \in S$.

We thus summarize as follows:

Proposition 1 *Let $S \equiv [\underline{\tau}, \overline{\tau}]^2$ with $(\underline{\tau}, \overline{\tau}) \in \Re^2$ and $0 < \underline{\tau} < \overline{\tau}$, and fix $(\tau_0^1, \tau_0^{*1}, \tau_0^2, \tau_0^{*2}) \in S^2$. For $i = 1, 2$, assume $\Delta W^i(\tau^1, \tau^{*1}, \tau^2, \tau^{*2}; \tau_0^i, \tau_0^{*i}) \cdot \Delta W^{*i}(\tau^1, \tau^{*1}, \tau^2, \tau^{*2}; \tau_0^i, \tau_0^{*i})$ is quasiconcave in $(\tau^i, \tau^{*i}) \in S$. Then there exists a vector $\tau_{hw} \equiv (\tau_{hw}^1, \tau_{hw}^{*1}, \tau_{hw}^2, \tau_{hw}^{*2}) \in S^2$ that is a Horn-Wolinsky solution.*

Notice that the proposition imposes quasiconcavity for the Nash Bargaining solution objective, $\Delta W^i(\tau^1, \tau^{*1}, \tau^2, \tau^{*2}; \tau_0^i, \tau_0^{*i}) \cdot \Delta W^{*i}(\tau^1, \tau^{*1}, \tau^2, \tau^{*2}; \tau_0^i, \tau_0^{*i})$, rather than for the individual welfare functions. We also take $(\tau_0^1, \tau_0^{*1}, \tau_0^2, \tau_0^{*2})$ to be an exogenous tariff vector. One possibility is that this vector corresponds to the prior or "standing" agreements in each bilateral relationship. Finally, we remark that the stated proposition establishes existence but does not ensure interiority. We are thus not yet in a position to characterize the Horn-Wolinsky solution using first-order conditions for optimization.

We now follow up on this latter point and establish sufficient conditions for an interior solution to the Horn-Wolinsky model, where an *interior Horn-Wolinsky solution* is a Horn-Wolinsky solution for which $(\tau_{hw}^1, \tau_{hw}^{*1}, \tau_{hw}^2, \tau_{hw}^{*2}) \in (\underline{\tau}, \overline{\tau})^4$.

Proposition 2 *Let $S \equiv [\underline{\tau}, \overline{\tau}]^2$ with $(\underline{\tau}, \overline{\tau}) \in \Re^2$ and $0 < \underline{\tau} < \overline{\tau}$. For any $(\tau^1, \tau^{*1}, \tau^2, \tau^{*2}) \in S^2$ and $(\tau_0^1, \tau_0^{*1}, \tau_0^2, \tau_0^{*2}) \in S^2$, and for $i = 1, 2$, assume $\Delta W^i(\tau^1, \tau^{*1}, \tau^2, \tau^{*2}; \tau_0^1, \tau_0^{*1}) \cdot \Delta W^{*i}(\tau^1, \tau^{*1}, \tau^2, \tau^{*2}; \tau_0^1, \tau_0^{*1})$ is strictly increasing in τ^i when $\tau^i = \underline{\tau}$, strictly decreasing in τ^i when $\tau^i = \overline{\tau}$, strictly increasing in τ^{*i} when $\tau^{*i} = \underline{\tau}$, and strictly decreasing in τ^{*i} when $\tau^{*i} = \overline{\tau}$. Then any Horn-Wolinsky solution must be an interior Horn-Wolinsky solution.*

Notice that these sufficient conditions are stronger than necessary, since we impose signs on boundary derivatives that hold regardless of the values assumed by other tariffs. Notice also that the conditions are imposed on the Nash Bargaining solution objective.

4 Discriminatory Tariffs: Sufficient Conditions for Pareto Gains

In this section, we suppose that an interior Horn-Wolinsky solution exists, and we establish a sense in which the resulting tariffs must be inefficient and too low. Specifically, we provide sufficient conditions under which it is possible to construct a particular perturbation where all countries gain by raising their tariffs.

To begin our analysis, we suppose that we have some bargaining parameter α and that the Horn-Wolinsky model under simultaneous bilateral bargaining delivers an outcome, $\tau_{hw} \equiv (\tau_{hw}^1, \tau_{hw}^{*1}, \tau_{hw}^2, \tau_{hw}^{*2})$, where τ_{hw} is an interior Horn-Wolinsky solution. As above, we represent the welfare of each country as a function of the vector of tariffs. Given interiority, we know that each tariff pair, $(\tau_{hw}^i, \tau_{hw}^{*i})$, is bilaterally efficient, holding fixed the other tariff pair. In other words, we know that our solution resides on the bilateral efficiency loci:

$$-\frac{\partial W}{\partial \tau^{*i}} / \frac{\partial W}{\partial \tau^i} = -\frac{\partial W^{*i}}{\partial \tau^{*i}} / \frac{\partial W^{*i}}{\partial \tau^i}, \text{ for } i = 1, 2. \tag{5}$$

In analogy with the assumptions in Bagwell-Staiger (2005) for points on the efficiency frontier, we assume that, at the Horn-Wolinsky solution tariff vector τ_{hw}, the welfare impacts of tariff changes satisfy the following restrictions: for $i, j = 1, 2$ and $i \neq j$,

$$\begin{aligned} \frac{\partial W}{\partial \tau^{i}} &> 0 \text{ and } \frac{\partial W^{*i}}{\partial \tau^{*i}} > 0 \\ \frac{\partial W}{\partial \tau^{*i}} &< 0 \text{ and } \frac{\partial W^{*i}}{\partial \tau^{i}} < 0 \\ \frac{\partial W^{*i}}{\partial \tau^{*j}} &> 0 \text{ and } \frac{\partial W^{*i}}{\partial \tau^{j}} > 0. \end{aligned} \tag{6}$$

Starting at any such Horn-Wolinsky solution as captured by (5), and under the assumptions given in (6), our claim now is that we can increase all four tariffs in a way that raises the welfare of all three countries. This directly suggests a local sense in which the Horn-Wolinsky tariffs are "too low" from an efficiency standpoint.

The idea of the perturbation draws from footnote 11 in Bagwell-Staiger (2005). Bagwell and Staiger (2005) consider an efficient tariff vector and suppose that the tangency condition in (5) holds between the home country and some foreign country $*i$. They then consider a two-step perturbation. In the first step, they increase τ^{i} and τ^{*i} in a fashion that maintains W^{*i}. This first-step perturbation results in no change in W^{*i}, a first-order increase in W^{*j} and a second-order loss in W (due to the tangency between the iso-welfare curves of the home country and foreign country $*i$). The second step is then to increase τ^{j} and decrease τ^{*j} in a fashion that maintains W^{*i}. This second-step perturbation results in no change in W^{*i}, a first-order loss in W^{*j} and a first-order gain in W. If the second-step perturbation is small relative to the first-step perturbation, then the Bagwell-Staiger perturbation in total results in no change in W^{*i} and first-order gains in W^{*j} and W, which contradicts the original hypothesis of an efficient tariff vector.

We want to consider here a similar perturbation, but there are three differences. First, we start with a situation in which the tangency condition (5) holds between the home country and both foreign countries. Second, we want to find a perturbation that generates welfare gains to each of the three countries. (By contrast, in the Bagwell-Staiger (2005) perturbation just defined, W^{*i} is unchanged.) Third, we want to find a perturbation under which all four tariffs are increased. (By contrast, in the single Bagwell-Staiger (2005) perturbation just described, τ^{*j} is decreased.)

The key idea is to do two Bagwell-Staiger (2005) perturbations simultaneously, so that each foreign country plays the role in the Bagwell-Staiger perturbation of "foreign country $*j$" in one perturbation and thus emerges with a welfare gain in the combined perturbation. If for each perturbation the second-step adjustment is small in comparison to the first-step adjustment, then the combined perturbation will also call for a higher tariff from each foreign country. In other words, we will construct a combined perturbation such that, for each foreign country, the first-step tariff increase that it undertakes when playing the role of foreign country $*i$ exceeds the second-step tariff decrease that it undertakes when playing the role of foreign country $*j$.

Starting at a tariff vector that satisfies (5), and under the assumption (6), we now consider the following perturbation:

$$d\tau^1 = d\tau^2 = \epsilon + \sigma \tag{7}$$

$$\begin{aligned} d\tau^{*1} &= (-\frac{\partial W^{*1}}{\partial \tau^1}/\frac{\partial W^{*1}}{\partial \tau^{*1}})\epsilon + (-\frac{\partial W^{*2}}{\partial \tau^1}/\frac{\partial W^{*2}}{\partial \tau^{*1}})\sigma \\ &= (-\frac{\partial W}{\partial \tau^1}/\frac{\partial W}{\partial \tau^{*1}})\epsilon + (-\frac{\partial W^{*2}}{\partial \tau^1}/\frac{\partial W^{*2}}{\partial \tau^{*1}})\sigma \end{aligned} \tag{8}$$

$$\begin{aligned} d\tau^{*2} &= (-\frac{\partial W^{*2}}{\partial \tau^2}/\frac{\partial W^{*2}}{\partial \tau^{*2}})\epsilon + (-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma \\ &= (-\frac{\partial W}{\partial \tau^2}/\frac{\partial W}{\partial \tau^{*2}})\epsilon + (-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma \end{aligned} \tag{9}$$

where the equalities in the second lines of (8) and (9) follow from the bilateral efficiency conditions (5) which the starting tariffs are assumed to satisfy, and where $\epsilon > 0$ and $\sigma > 0$ are both small. We give a further condition below concerning the relative magnitudes of ϵ and σ.

We can now compute the welfare differentials. For the home country, we get

$$\begin{aligned} dW &= (\frac{\partial W}{\partial \tau^1} + \frac{\partial W}{\partial \tau^2})(\epsilon + \sigma) + \frac{\partial W}{\partial \tau^{*1}}[(-\frac{\partial W}{\partial \tau^1}/\frac{\partial W}{\partial \tau^{*1}})\epsilon + (-\frac{\partial W^{*2}}{\partial \tau^1}/\frac{\partial W^{*2}}{\partial \tau^{*1}})\sigma] \\ &\quad + \frac{\partial W}{\partial \tau^{*2}}[(-\frac{\partial W}{\partial \tau^2}/\frac{\partial W}{\partial \tau^{*2}})\epsilon + (-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma]. \end{aligned}$$

Thus,

$$dW = [\frac{\partial W}{\partial \tau^1} + \frac{\partial W}{\partial \tau^2} + \frac{\partial W}{\partial \tau^{*1}}(-\frac{\partial W^{*2}}{\partial \tau^1}/\frac{\partial W^{*2}}{\partial \tau^{*1}}) + \frac{\partial W}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})]\sigma > 0,$$

where the inequality follows since $\sigma > 0$ and (6) holds at the original tariff vector.

For foreign country $*1$, we get

$$\begin{aligned} dW^{*1} &= (\frac{\partial W^{*1}}{\partial \tau^1} + \frac{\partial W^{*1}}{\partial \tau^2})(\epsilon + \sigma) + \frac{\partial W^{*1}}{\partial \tau^{*1}}[(-\frac{\partial W^{*1}}{\partial \tau^1}/\frac{\partial W^{*1}}{\partial \tau^{*1}})\epsilon + (-\frac{\partial W^{*2}}{\partial \tau^1}/\frac{\partial W^{*2}}{\partial \tau^{*1}})\sigma] \\ &\quad + \frac{\partial W^{*1}}{\partial \tau^{*2}}[(-\frac{\partial W^{*2}}{\partial \tau^2}/\frac{\partial W^{*2}}{\partial \tau^{*2}})\epsilon + (-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma] \end{aligned}$$

Thus,

$$dW^{*1} = [\frac{\partial W^{*1}}{\partial \tau^2} + \frac{\partial W^{*1}}{\partial \tau^{*2}}(-\frac{\partial W^{*2}}{\partial \tau^2}/\frac{\partial W^{*2}}{\partial \tau^{*2}})]\epsilon + [\frac{\partial W^{*1}}{\partial \tau^1} + \frac{\partial W^{*1}}{\partial \tau^{*1}}(-\frac{\partial W^{*2}}{\partial \tau^1}/\frac{\partial W^{*2}}{\partial \tau^{*1}})]\sigma$$

As noted above, we can think of the perturbation here as a combination of two Bagwell-Staiger (2005) perturbations, which we might think of as Home-$*1$ and Home-$*2$ perturbations (with the designated foreign country playing the role of foreign country $*i$ in Bagwell and Staiger, 2005). The ϵ part of dW^{*1} is then the gain in W^{*1} from the Home-$*2$ step-1 increase in τ^2 and τ^{*2}, where there is no first-order effect on W^{*1} from the Home-$*1$ step-1 increase in τ^1 and τ^{*1}. The σ part of

dW^{*1} is then the loss in W^{*1} from the Home-$*2$ step-2 increase in τ^1 and decrease in τ^{*1} to keep W^{*2} fixed, where by construction there is no effect on W^{*1} from the Home-$*1$ step-2 increase in τ^2 and decrease in τ^{*2} that keeps W^{*1} fixed.

Under our assumption that the initial tariff vector satisfies (6), the term in dW^{*1} that is multiplied by ϵ is positive while the term that is multiplied by σ is negative; therefore, if ϵ is large relative to σ in the specific sense that

$$\epsilon > \{\frac{-[\frac{\partial W^{*1}}{\partial \tau^1} + \frac{\partial W^{*1}}{\partial \tau^{*1}}(-\frac{\partial W^{*2}}{\partial \tau^1}/\frac{\partial W^{*2}}{\partial \tau^{*1}})]}{\frac{\partial W^{*1}}{\partial \tau^2} + \frac{\partial W^{*1}}{\partial \tau^{*2}}(-\frac{\partial W^{*2}}{\partial \tau^2}/\frac{\partial W^{*2}}{\partial \tau^{*2}})}\}\sigma$$

then $dW^{*1} > 0$. An exactly symmetric argument holds for foreign country $*2$.

Allowing for $i = 1, 2$, we thus select $\epsilon > 0$ and $\sigma > 0$ such that

$$\epsilon > \max_{i,j=1,2,i\neq j}\{\frac{-[\frac{\partial W^{*i}}{\partial \tau^i} + \frac{\partial W^{*i}}{\partial \tau^{*i}}(-\frac{\partial W^{*j}}{\partial \tau^i}/\frac{\partial W^{*j}}{\partial \tau^{*i}})]}{\frac{\partial W^{*i}}{\partial \tau^j} + \frac{\partial W^{*i}}{\partial \tau^{*j}}(-\frac{\partial W^{*j}}{\partial \tau^j}/\frac{\partial W^{*j}}{\partial \tau^{*j}})}\}\sigma \tag{10}$$

Under (10), we may conclude that the perturbation raises the welfare of each country.

The remaining issue is to confirm that the perturbation increases each tariff. It is clear from (7) that $d\tau^1 = d\tau^2 > 0$. Referring to (8) and (9), we see for $d\tau^{*i}$ that the coefficient on ϵ is positive while that on σ is negative. Thus, we have that $d\tau^{*i} > 0$ for $i, j = 1, 2, i \neq j$ if and only if

$$\epsilon > \max_{i,j=1,2,i\neq j}[\frac{-\frac{\partial W^{*j}}{\partial \tau^i}/\frac{\partial W^{*j}}{\partial \tau^{*i}}}{\frac{\partial W^{*i}}{\partial \tau^i}/\frac{\partial W^{*i}}{\partial \tau^{*i}}}]\sigma. \tag{11}$$

We now have

Proposition 3 *For a given bargaining parameter α, suppose the Horn-Wolinsky model under simultaneous bilateral bargaining delivers an interior Horn-Wolinsky solution, $\tau_{hw} \equiv (\tau^1_{hw}, \tau^{*1}_{hw}, \tau^2_{hw}, \tau^{*2}_{hw})$, captured by (5). Suppose at this tariff vector that (6) holds. For sufficiently small $\epsilon > 0$, $\sigma > 0$ satisfying (10) and (11), the perturbation defined in (7)-(9) raises all tariffs and generates welfare gains for all three countries.*

We note that welfare gains accrue to all countries without separately assuming that (11) holds. The role of (11) is simply to ensure that all tariffs are increased as part of the perturbation.[13]

We now have established that conditions exist under which, starting at any interior Horn-Wolinsky solution, all three countries can gain through a perturbation under which they all raise their tariffs. We thus have formalized an interpretation in which tariffs are inefficient in the sense

[13] It can also be shown that (10) implies (11) under additional assumptions. For example, given our assumptions in (6), this implication holds in a symmetric setting under a natural further assumption. A setting is symmetric if foreign countries $*1$ and $*2$ have symmetric welfare functions W^{*1} and W^{*2} and if tariffs are symmetric with $\tau^1 = \tau^2$ and $\tau^{*1} = \tau^{*2}$. For a symmetric setting, a natural further assumption is that $\partial W^{*i}/\partial \tau^i + \partial W^{*i}/\partial \tau^j < 0$, so that for each foreign country $*i$ the cost of facing a higher home-country tariff exceeds the benefit of the other foreign country facing the higher home-country tariff. In Section 6, we explicitly impose this further assumption in (16) in the context of our analysis of MFN tariffs.

of being too low, at any interior Horn-Wolinsky solution. This result complements our work in the previous section, where we provide conditions under which an interior Horn-Wolinsky solution satisfying the tangency conditions in (5) exists.

5 Discriminatory Tariffs: Necessary Conditions for Pareto Gains

Our results in the preceding section may be understood as providing sufficient conditions for Pareto gains through tariff increases; specifically, starting at an interior Horn-Wolinsky solution, we construct a particular perturbation under which all countries gain by raising their tariffs. In this section, we again start at an interior Horn-Wolinsky solution, but we now examine the necessary conditions for perturbations that give Pareto gains. Our main finding is that, starting at an interior Horn-Wolinsky solution, if all countries enjoy weak welfare gains under a perturbation, then the perturbation cannot be characterized by "opportunistic" bilateral tariff changes in both bilateral relationships. We also show that, if at least one country strictly gains under such a perturbation, then there must exist a bilateral relationship in which at least one tariff rises while the other tariff in that relationship can fall but only to a limited extent. In this sense, our findings in this section reinforce the interpretation we formalize in the previous section that, at any interior Horn-Wolinsky solution, tariffs are inefficiently low.

To formalize these arguments, we begin with some definitions. Let $\tau_{hw} \equiv (\tau^1_{hw}, \tau^{*1}_{hw}, \tau^2_{hw}, \tau^{*2}_{hw})$ denote an interior Horn-Wolinsky solution, where we assume again that (5) and (6) hold at this vector. Starting at τ_{hw}, we consider a perturbation $d\boldsymbol{\tau} \equiv (d\tau^1, d\tau^{*1}, d\tau^2, d\tau^{*2})$. It is convenient to decompose the perturbation into the bilateral tariff changes that are implied for each bilateral relationship, $(d\tau^1, d\tau^{*1})$ and $(d\tau^2, d\tau^{*2})$. For $i = 1, 2$, it is also convenient to define for the Home-$*i$ bilateral relationship a function $\widehat{\tau}^i$ that maps the tariff of foreign country $*i$ to the tariff that the home country applies on exports from foreign country $*i$. Our starting point is an interior Horn-Wolinsky solution, and so we assume that the function captures this solution: $\tau^i_{hw} = \widehat{\tau}^i(\tau^{*i}_{hw})$. To ensure that the function $\widehat{\tau}^i$ also captures the perturbation as it relates to the Home-$*i$ bilateral relationship, we require further that $d\tau^i = [d\widehat{\tau}^i(\tau^{*i}_{hw})/d\tau^{*i}]d\tau^{*i}$. We can then represent the perturbation as changes in foreign tariffs, $d\tau^{*1}$ and $d\tau^{*2}$, with the corresponding changes in home tariffs captured as $d\tau^1 = [d\widehat{\tau}^1(\tau^{*1}_{hw})/d\tau^{*1}]d\tau^{*1}$ and $d\tau^2 = [d\widehat{\tau}^2(\tau^{*2}_{hw})/d\tau^{*2}]d\tau^{*2}$. Thus, for a given perturbation, the bilateral tariff changes in the Home-$*i$ bilateral relationship can be represented as $(d\tau^i, d\tau^{*i})$ where $d\tau^i = [d\widehat{\tau}^i(\tau^{*i}_{hw})/d\tau^{*i}]d\tau^{*i}$.

For $i, j = 1, 2$ with $i \neq j$, we now say that the perturbation entails an *opportunistic bilateral tariff change in the Home-$*j$ bilateral relationship* if the bilateral tariff change described by $(d\tau^j, d\tau^{*j})$ reduces the welfare of foreign country $*i$:

$$[\frac{\partial W^{*i}}{\partial \tau^{*j}} + \frac{\partial W^{*i}}{\partial \tau^j}\frac{d\widehat{\tau}^j(\tau^{*j}_{hw})}{d\tau^{*j}}]d\tau^{*j} < 0, \tag{12}$$

where $d\tau^{*j} \neq 0$ thus holds given (12). As a general matter, we note that an opportunistic bilateral tariff change in the Home-$*j$ bilateral relationship does not necessarily imply that the perturbation $d\boldsymbol{\tau}$ reduces the welfare of foreign country $*i$, since the perturbation includes as well the tariff changes $(d\tau^i, d\tau^{*i})$ in the Home-$*i$ bilateral relationship.

Let us now consider a perturbation that entails an opportunistic bilateral tariff change in both bilateral relationships. In other words, we consider now a perturbation for which (12) holds for $i, j = 1, 2$ and $i \neq j$. Each foreign country then suffers from the tariff changes that occur in the "other" bilateral relationship. We ask the following question: Starting at an interior Horn-Wolinsky solution, is it possible that such a perturbation can generate weak welfare gains for all countries? We argue next that the answer to this question is "no," from which it follows that a perturbation generating weak welfare gains for all countries necessarily has non-opportunistic bilateral tariff changes for at least one bilateral relationship.

To make this argument, let us suppose to the contrary that the perturbation satisfies (12) for $i, j = 1, 2$ with $i \neq j$ and yet generates weak welfare gains for all three countries. Consider now the welfare change under the perturbation for foreign country $*i$:

$$dW^{*i} = [\frac{\partial W^{*i}}{\partial \tau^{*j}} + \frac{\partial W^{*i}}{\partial \tau^j}\frac{d\widehat{\tau}^j(\tau^{*j}_{hw})}{d\tau^{*j}}]d\tau^{*j} + [\frac{\partial W^{*i}}{\partial \tau^{*i}} + \frac{\partial W^{*i}}{\partial \tau^i}\frac{d\widehat{\tau}^i(\tau^{*i}_{hw})}{d\tau^{*i}}]d\tau^{*i} \geq 0,$$

where the inequality follows from the assumption that the welfare change is non-negative for all countries. Under (12), we see that the first term in this expression is negative; thus, it follows that

$$[\frac{\partial W^{*i}}{\partial \tau^{*i}} + \frac{\partial W^{*i}}{\partial \tau^i}\frac{d\widehat{\tau}^i(\tau^{*i}_{hw})}{d\tau^{*i}}]d\tau^{*i} > 0.$$

Using $\partial W^{*i}/\partial \tau^i < 0$ under (6), we may rewrite this inequality equivalently as

$$[\frac{\partial W^{*i}}{\partial \tau^{*i}}/\frac{\partial W^{*i}}{\partial \tau^i} + \frac{d\widehat{\tau}^i(\tau^{*i}_{hw})}{d\tau^{*i}}]d\tau^{*i} < 0, \tag{13}$$

where the inequality in (13) holds for $i, j = 1, 2$ with $i \neq j$.

We consider next the welfare change under the perturbation for the home country. We find that

$$\begin{aligned} dW &= [\frac{\partial W}{\partial \tau^{*j}} + \frac{\partial W}{\partial \tau^j}\frac{d\widehat{\tau}^j(\tau^{*j}_{hw})}{d\tau^{*j}}]d\tau^{*j} + [\frac{\partial W}{\partial \tau^{*i}} + \frac{\partial W}{\partial \tau^i}\frac{d\widehat{\tau}^i(\tau^{*i}_{hw})}{d\tau^{*i}}]d\tau^{*i} \\ &= [\frac{\partial W}{\partial \tau^{*j}}/\frac{\partial W}{\partial \tau^j} + \frac{d\widehat{\tau}^j(\tau^{*j}_{hw})}{d\tau^{*j}}]d\tau^{*j}\frac{\partial W}{\partial \tau^j} + [\frac{\partial W}{\partial \tau^{*i}}/\frac{\partial W}{\partial \tau^i} + \frac{d\widehat{\tau}^i(\tau^{*i}_{hw})}{d\tau^{*i}}]d\tau^{*i}\frac{\partial W}{\partial \tau^i}, \end{aligned}$$

where we use $\partial W/\partial \tau^i > 0$ for $i = 1, 2$ by (6). We now use the fact that an interior Horn-Wolinsky solution is bilaterally efficient and thus characterized by a tangency. In particular, using (5), we now have that

$$dW = \{[\frac{\partial W^{*j}}{\partial \tau^{*j}}/\frac{\partial W^{*j}}{\partial \tau^j} + \frac{d\widehat{\tau}^j(\tau^{*j}_{hw})}{d\tau^{*j}}]d\tau^{*j}\}\frac{\partial W}{\partial \tau^j} + \{[\frac{\partial W^{*i}}{\partial \tau^{*i}}/\frac{\partial W^{*i}}{\partial \tau^i} + \frac{d\widehat{\tau}^i(\tau^{*i}_{hw})}{d\tau^{*i}}]d\tau^{*i}\}\frac{\partial W}{\partial \tau^i} < 0,$$

where the inequality follows since each term in curly brackets is negative by (13) and $\partial W/\partial \tau^i > 0$

for $i = 1, 2$ by (6). Finally, we note that $dW < 0$ is a contradiction to our assumption that the perturbation generates weak welfare gains for all countries.

The following proposition summarizes our finding.

Proposition 4 *For a given bargaining parameter α, suppose the Horn-Wolinsky model under simultaneous bilateral bargaining delivers an interior Horn-Wolinsky solution, $\tau_{hw} \equiv (\tau^1_{hw}, \tau^{*1}_{hw}, \tau^2_{hw}, \tau^{*2}_{hw})$, captured by (5). Suppose at this tariff vector that (6) holds. Starting at this solution, a small perturbation $d\boldsymbol{\tau} \equiv (d\tau^1, d\tau^{*1}, d\tau^2, d\tau^{*2})$ generates weak welfare gains for all three countries only if the bilateral tariff change in at least one bilateral relationship is not opportunistic.*

Intuitively, if a perturbation from an interior Horn-Wolinsky solution entails opportunistic bilateral tariff changes for the Home-$*i$ bilateral relationship, then foreign country $*j$ can enjoy a weak gain under the perturbation only if it gains from the bilateral tariff changes in the Home-$*j$ bilateral relationship. For an interior Horn-Wolinsky solution, however, we know that the bilateral tariffs in the Home-$*j$ bilateral relationship are set in a bilaterally efficient manner; thus, foreign country $*j$ can gain from a change in the bilateral tariffs that it and the home country apply to each other only if the home country loses from this change. Continuing from here, if the home country is to enjoy a weak gain from the perturbation, then its loss in the Home-$*j$ bilateral relationship must be offset by a gain in the Home-$*i$ bilateral relationship. But by analogous reasoning, if the interior Horn-Wolinsky solution entails opportunistic bilateral tariff changes for the Home-$*j$ bilateral relationship, then foreign country $*i$ can enjoy a weak gain from the perturbation only if it, too, enjoys a gain in the Home-$*i$ bilateral relationship. Since the bilateral tariffs in the Home-$*i$ bilateral relationship are likewise set in a bilaterally efficient manner, however, it is not possible to find bilateral tariff changes for the Home-$*i$ bilateral relationship such that both the home country and foreign country $*i$ enjoy gains.

We conclude this section by exploring the implications of this proposition for the nature of the underlying tariff changes that a Pareto-improving perturbation must deliver. To develop our findings, we return to the trade model. We note first that foreign country $*i$ experiences a welfare change as a consequence of bilateral tariff changes in the Home-$*j$ bilateral relationship if and only if the bilateral tariff changes alter foreign country $*i$'s terms of trade, $\widetilde{p}^{wi}$. We now confirm that, under our existing assumptions, starting at an interior Horn-Wolinsky solution, foreign country $*i$ suffers a welfare loss when it faces an externally generated deterioration in its terms of trade; that is, we show that, at the interior Horn-Wolinsky solution tariff vector τ_{hw}, and for $i = 1, 2$,

$$\frac{d}{d\widetilde{p}^{wi}} w^{*i}(p^{*i}, \widetilde{p}^{wi}) > 0, \tag{14}$$

where we recall that $p^{*i} = (1/\tau^{*i})\widetilde{p}^{wi}$. To confirm that (14) necessarily holds under our existing assumptions, we observe from (6) that at the interior Horn-Wolinsky solution tariff vector τ_{hw}, and for $i = 1, 2$,

$$\frac{\partial W^{*i}}{\partial \tau^i} = [\frac{d}{d\widetilde{p}^{wi}} w^{*i}(p^{*i}, \widetilde{p}^{wi})][\frac{d\widetilde{p}^{wi}}{d\tau^i}] < 0.$$

Since $\widetilde{p}^{wi}$ is decreasing in τ^i, it is now evident that (14) holds.

Given (14) and for $i, j = 1, 2$ with $i \neq j$, we see that a perturbation entails an *opportunistic bilateral tariff change in the Home-$*j$ bilateral relationship* if and only if the bilateral tariff change described by $(d\tau^j, d\tau^{*j})$ generates a terms-of-trade loss for foreign country $*i$. Accordingly, by (14), the meaning of our proposition is that weak Pareto welfare gains are possible under a perturbation only if the bilateral tariff changes for at least one bilateral relationship result in a weak terms-of-trade gain for the foreign country that is not a member of that relationship.

We are now ready to explore the necessary features of the tariff changes that a Pareto-improving perturbation must deliver. Specifically, we consider a perturbation with two properties: it generates weak welfare gains for all countries, and at least one country actually gains under the perturbation. The latter property rules out the trivial possibility where all tariffs are unaltered. As just noted, the first property ensures that, for some $i, j = 1, 2$ with $i \neq j$, the bilateral tariff changes in the Home-$*j$ bilateral relationship generates a weak terms-of-trade gain for foreign country $*i$.

Consider such a perturbation. A first point is that the perturbation must entail a change in τ^j, τ^{*j} or both. To see why, suppose that the perturbation changes neither τ^j nor τ^{*j}. Given that, at an interior Horn-Wolinsky solution, τ^i and τ^{*i} are set in a bilaterally efficient manner for the Home-$*i$ bilateral relationship, any change in τ^i and τ^{*i} must result in a welfare loss for the home country or foreign country $*i$. Furthermore, if τ^i and τ^{*i} were also unaltered, then the perturbation would fail to generate an actual welfare gain for any country. We conclude that the assumed properties of the perturbation necessitate a change in τ^j and/or τ^{*j}. Since $\widetilde{p}^{wi}$ is increasing in τ^j and τ^{*j}, we may further observe that foreign country $*i$ cannot enjoy the assumed weak terms-of-trade gain if τ^j and τ^{*j} are both reduced. It follows that τ^j rises, τ^{*j} rises, or both τ^j and τ^{*j} rise.

We may thus conclude the section with the following proposition:

Proposition 5 *For a given bargaining parameter α, suppose the Horn-Wolinsky model under simultaneous bilateral bargaining delivers an interior Horn-Wolinsky solution, $\tau_{hw} \equiv (\tau^1_{hw}, \tau^{*1}_{hw}, \tau^2_{hw}, \tau^{*2}_{hw})$, captured by (5). Suppose at this tariff vector that (6) holds. Starting at this solution, a small perturbation $d\boldsymbol{\tau} \equiv (d\tau^1, d\tau^{*1}, d\tau^2, d\tau^{*2})$ generates a welfare gain for at least one country while not lowering the welfare of any other country only if at least one tariff rises.*

The proposition establishes that the described Pareto improvement requires an increase in at least one tariff, but it is important to recognize that the underlying argument also places restrictions on the extent to which other tariffs can fall. In particular, given (14), we know that a weak Pareto improvement requires that, in at least one bilateral relationship, the associated bilateral tariff changes generate a weak terms-of-trade gain for the non-member foreign country. As we argue above, if we assume further that the perturbation generates an actual welfare gain for at least one country, then we can conclude that at least one tariff in this bilateral relationship actually rises. The other tariff in this bilateral relationship may rise as well or it could fall. But if it falls, it cannot fall to such an extent as to reverse the weak terms-of-trade gain that the non-member foreign country must enjoy.

6 MFN Tariffs: Sufficient Conditions for Pareto Gains

We now suppose that the home-country tariffs satisfy the MFN rule, so that $\tau \equiv \tau^1 = \tau^2$. We assume for this setting that the home country engages in a bilateral bargain with only its principal supplier, which we take to be foreign country $*1$ for simplicity. The tariff that foreign country $*2$ applies is thus left untouched and remains fixed at some exogenous level, $\overline{\tau}^{*2}$. Our goal in this section is to define the Horn-Wolinsky MFN solution for this setting and then identify conditions under which it is inefficient when efficiency is evaluated relative to the full space of tariff policies, $\boldsymbol{\tau} \equiv (\tau^1, \tau^{*1}, \tau^2, \tau^{*2})$. We also show that, in one important case, all countries can gain starting at the Horn-Wolinsky MFN solution when all four tariffs are reduced in an appropriate fashion. For this case, we thus formalize an interpretation in which tariffs are inefficient in the sense of being too high at a Horn-Wolinsky MFN solution. More generally, our findings in this section indicate that the MFN rule works as a partial counterbalance to the forces that arise when discriminatory tariffs are permitted and which lead tariffs to be inefficiently low at any interior Horn-Wolinsky solution.

Formally, with the home-country tariff satisfying the MFN rule so that $\tau \equiv \tau^1 = \tau^2$, and with foreign country $*2$'s tariff fixed at an exogenous level, $\overline{\tau}^{*2}$, we assume that the negotiation between the home country and foreign country $*1$ is captured by a Nash bargaining solution, with bargaining parameter α. We refer to the associated outcome as the *Horn-Wolinsky MFN solution,* and we represent this outcome with the tariff vector $\widetilde{\boldsymbol{\tau}}_{hw} \equiv (\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw}, \widetilde{\tau}_{hw}, \overline{\tau}^{*2})$, where we have used that the MFN rule gives $\widetilde{\tau}^{*1}_{hw} = \widetilde{\tau}^{*2}_{hw} \equiv \widetilde{\tau}_{hw}$ and that the tariff of foreign country $*2$ is fixed so that $\widetilde{\tau}^{*2}_{hw} \equiv \overline{\tau}^{*2}$. Assuming an interior solution, the resulting bargaining tariff pair, $(\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw})$, is bilaterally efficient, holding fixed τ^{*2} at $\overline{\tau}^{*2}$. In other words, the solution resides on the bilateral efficiency loci for the MFN setting:

$$-\frac{\partial W}{\partial \tau^{*1}} / \frac{dW}{d\tau} = -\frac{\partial W^{*1}}{\partial \tau^{*1}} / \frac{dW^{*1}}{d\tau}, \tag{15}$$

where the total derivatives reflect the fact that welfare functions are expressed as functions of the tariff vector $\boldsymbol{\tau} \equiv (\tau^1, \tau^{*1}, \tau^2, \tau^{*2})$ where in the MFN setting $\tau \equiv \tau^1 = \tau^2$.

In analogy with the assumptions in Bagwell-Staiger (2005) for points on the MFN-efficiency frontier, we assume that, at the Horn-Wolinsky MFN solution tariff satisfying (15) and given $\tau^{*2} \equiv \overline{\tau}^{*2}$, the welfare impacts of tariff changes for $i, j = 1, 2$ and $i \neq j$ satisfy (6) and also

$$\frac{dW^{*i}}{\partial \tau} < 0. \tag{16}$$

Starting at the Horn-Wolinsky MFN solution as captured by (15), and under the assumptions given in (6) and (16), we seek to identify conditions under which the Horn-Wolinsky MFN solution is inefficient relative to the full space of tariff policies, $\boldsymbol{\tau} \equiv (\tau^1, \tau^{*1}, \tau^2, \tau^{*2})$. Our analysis is organized around cases. As noted above, we identify one case under which, starting at the Horn-Wolinsky MFN solution, it is possible to reduce all tariffs in a fashion that generates Pareto gains.

Case 1 To establish our results, we must distinguish between three cases. The first case captures the possibility that, at the Horn-Wolinsky MFN solution and the induced MFN world price $\widetilde{p}^w(\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw}, \overline{\tau}^{*2})$, the home country would prefer more trade whereas foreign country $*1$ would prefer less trade. This is the case for which welfare gains accrue to all countries when all tariffs are reduced in an appropriate fashion.

The idea of the perturbation is as follows. In step 1 of the perturbation, the home country and foreign country $*1$ lower τ and τ^{*1} in a fashion that maintains W^{*1}. Under the first case, this change necessitates a terms-of-trade improvement for foreign country $*1$ (i.e., an increase in $\widetilde{p}^w$) as compensation for the increased trade volume. Due to the tangency condition (15), the home country then experiences only a second-order loss from this first-step adjustment. By contrast, foreign country $*2$ has the same terms-of-trade as foreign country $*1$ and thus enjoys a terms-of-trade gain, and thus a first-order welfare gain, from the first-step adjustment. In the second step of the perturbation, we utilize the full tariff space and raise τ^2 while lowering τ^{*2}, again so as to maintain W^{*1}. This second change generates a first-order gain for the home country and a first-order loss for foreign country $*2$. If the second-step adjustment is sufficiently small in magnitude, however, then the home country and foreign country $*2$ enjoy first-order gains overall and all tariffs drop. At this point, we could allow a very small increase in the tariff of foreign country $*1$ to ensure that all countries gain from a perturbation in which all tariffs are reduced.

We now formalize this idea. We begin by expressing the conditions that define our first case:

$$-\frac{\partial W}{\partial \tau^{*1}} / \frac{dW}{d\tau} = -\frac{\partial W^{*1}}{\partial \tau^{*1}} / \frac{dW^{*1}}{d\tau} > -\frac{\partial W^{*2}}{\partial \tau^{*1}} / \frac{dW^{*2}}{d\tau}, \tag{17}$$

where the expressions are evaluated at the Horn-Wolinsky MFN solution and where total derivatives again reflect the fact that $\tau \equiv \tau^1 = \tau^2$ in the MFN setting. The equality in (17) captures a tangency between the iso-welfare curves of the home country and foreign country $*1$ in a graph with τ on the y axis and τ^{*1} on the x axis, which we illustrate in Figure 1. The inequality indicates that the iso-world-price curve, which gives combinations of τ and τ^{*1} that preserve $\widetilde{p}^w(\tau, \tau^{*1}, \tau^{*2})$ at the value obtained at the Horn-Wolinsky MFN solution, is flatter, as Figure 1 depicts. This ensures that, at the prevailing world price, the home country desires more trade whereas foreign country $*1$ prefers less trade. Finally, since foreign country $*2$ is impacted by changes in τ and τ^{*1} only insofar as those changes impact the world price, we may equivalently express the slope of the iso-world-price curve in terms of the slope of the iso-welfare curve for foreign country $*2$.

Starting at a tariff vector that satisfies (15) and (17), and under the assumptions (6) and (16), we now consider the following perturbation:

$$d\tau^1 = -\epsilon \tag{18}$$

$$d\tau^2 = -\epsilon + \sigma \tag{19}$$

$$d\tau^{*1} = (-\frac{dW}{d\tau} / \frac{\partial W}{\partial \tau^{*1}})(-\epsilon) = (-\frac{dW^{*1}}{d\tau} / \frac{\partial W^{*1}}{\partial \tau^{*1}})(-\epsilon) \tag{20}$$

$$d\tau^{*2} = (-\frac{\partial W^{*1}}{\partial \tau^2} / \frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma \tag{21}$$

where the equality in (20) follows from the bilateral efficiency condition (15) which the starting tariffs are assumed to satisfy, and where $\epsilon > 0$ and $\sigma > 0$ are both small. We give a further condition below concerning the relative magnitudes of ϵ and σ.

We can now compute the welfare differentials. For the home country, we get

$$dW = \frac{dW}{d\tau}(-\epsilon) + \frac{\partial W}{\partial \tau^2}\sigma + \frac{\partial W}{\partial \tau^{*1}}(-\frac{dW}{d\tau} / \frac{\partial W}{\partial \tau^{*1}})(-\epsilon) + \frac{\partial W}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2} / \frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma.$$

Thus,

$$dW = [\frac{\partial W}{\partial \tau^2} + \frac{\partial W}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2} / \frac{\partial W^{*1}}{\partial \tau^{*2}})]\sigma > 0,$$

where the inequality follows since $\sigma > 0$ and (6) holds at the original tariff vector.

For foreign country $*1$, we get

$$dW^{*1} = \frac{dW^{*1}}{d\tau}(-\epsilon) + \frac{\partial W^{*1}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau} / \frac{\partial W^{*1}}{\partial \tau^{*1}})(-\epsilon) + \frac{\partial W^{*1}}{\partial \tau^2}\sigma + \frac{\partial W^{*1}}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2} / \frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma = 0,$$

which confirms that the perturbation maintains the welfare of foreign country $*1$.

Finally, for foreign country $*2$, we get

$$\begin{aligned} dW^{*2} &= \frac{dW^{*2}}{d\tau}(-\epsilon) + \frac{\partial W^{*2}}{\partial \tau^2}\sigma + \frac{\partial W^{*2}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau} / \frac{\partial W^{*1}}{\partial \tau^{*1}})(-\epsilon) + \frac{\partial W^{*2}}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2} / \frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma \\ &= [\frac{dW^{*2}}{d\tau} + \frac{\partial W^{*2}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau} / \frac{\partial W^{*1}}{\partial \tau^{*1}})](-\epsilon) + [\frac{\partial W^{*2}}{\partial \tau^2} - \frac{\partial W^{*1}}{\partial \tau^2}]\sigma. \end{aligned}$$

For the final expression presented here, we know from (6) that the bracketed expression in the second term is negative. Our next task is to use (6), (16) and (17) to sign the bracketed expression in the first term.

To this end, we recall the inequality in (17):

$$-\frac{\partial W^{*1}}{\partial \tau^{*1}} / \frac{dW^{*1}}{d\tau} > -\frac{\partial W^{*2}}{\partial \tau^{*1}} / \frac{dW^{*2}}{d\tau}.$$

Using (6) and (16), we can re-write this inequality as

$$\frac{dW^{*2}}{d\tau} + \frac{\partial W^{*2}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau} / \frac{\partial W^{*1}}{\partial \tau^{*1}}) < 0, \tag{22}$$

indicating that the bracketed expression in the first term to which we refer above is also negative.

It now follows that $dW^{*2} > 0$ if and only if

$$\epsilon > \{\frac{\frac{\partial W^{*2}}{\partial \tau^2} - \frac{\partial W^{*1}}{\partial \tau^2}}{\frac{dW^{*2}}{d\tau} + \frac{\partial W^{*2}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau} / \frac{\partial W^{*1}}{\partial \tau^{*1}})}\}\sigma. \tag{23}$$

Thus, since the expression in braces is positive, $dW^{*2} > 0$ holds if the second-step perturbation parameterized by σ is small relative to the first-step perturbation parameterized by ϵ.

Consider now the direction of the tariff changes. Given (6) and (16), it is straightforward to see from (18)-(21) that all tariffs are reduced if $\epsilon > \sigma$. But using (6), (16) and (22), it is also straightforward to show that the expression in braces in (23) is greater than 1, and hence $\epsilon > \sigma$ is assured by (23).

A final point is that we can ensure welfare gains for all countries, including foreign country $*1$, if we augment the perturbation to include an arbitrarily small increase in τ^{*1}. This augmented perturbation raises all three welfare levels while reducing all four tariffs.

We now have:

Proposition 6 *For a given bargaining parameter α, suppose the Horn-Wolinsky MFN solution delivers an interior solution, $\widetilde{\boldsymbol{\tau}}_{hw} \equiv (\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw}, \widetilde{\tau}_{hw}, \overline{\tau}^{*2})$, as captured by (15). Suppose at this tariff vector that (6) and (16) hold. Consider case 1 as defined by (17). For sufficiently small $\epsilon > 0$, $\sigma > 0$ satisfying (23), the perturbation defined in (18)-(21) when augmented with an arbitrarily small increase in τ^{*1} has the following effects: all tariffs are reduced, and all countries enjoy welfare gains.*

Case 2 The second case captures the possibility that, at the Horn-Wolinsky MFN solution and the induced MFN world price $\widetilde{p}^w(\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw}, \overline{\tau}^{*2})$, the home country would prefer less trade whereas foreign country $*1$ would prefer more trade.

The idea of the perturbation is as follows. In step 1 of the perturbation, the home country and foreign country $*1$ raise τ and τ^{*1} in a fashion that maintains W^{*1}. Under the second case, this change necessitates a terms-of-trade improvement for foreign country $*1$ (i.e., an increase in $\widetilde{p}^w$) as compensation for the reduced trade volume. Due to the tangency condition (15), the home country then experiences only a second-order loss from this first-step adjustment. By contrast, foreign country $*2$ has the same terms-of-trade as foreign country $*1$ and thus enjoys a terms-of-trade gain, and thus a first-order welfare gain, from the first-step adjustment. In the second step of the perturbation, we utilize the full tariff space and raise τ^2 while lowering τ^{*2}, again so as to maintain W^{*1}. This second change generates a first-order gain for the home country and a first-order loss for foreign country $*2$. If the second-step adjustment is sufficiently small in magnitude, however, then the home country and foreign country $*2$ enjoy first-order gains overall. At this point, we could allow a very small increase in the tariff of foreign country $*1$ to ensure that all countries gain from the perturbation.

One difference between the first and second cases is that not all tariffs move in the same direction in the perturbation that we use for the second case. In particular, in the perturbation that we use for the second case, τ^1, τ^2 and τ^{*1} rise while τ^{*2} falls.

To develop our formal analysis, we begin by expressing the conditions that define our second case:

$$-\frac{\partial W}{\partial \tau^{*1}} / \frac{dW}{d\tau} = -\frac{\partial W^{*1}}{\partial \tau^{*1}} / \frac{dW^{*1}}{d\tau} < -\frac{\partial W^{*2}}{\partial \tau^{*1}} / \frac{dW^{*2}}{d\tau}, \tag{24}$$

where the expressions are evaluated at the Horn-Wolinsky MFN solution and where total derivatives again reflect the fact that $\tau \equiv \tau^1 = \tau^2$ in the MFN setting. As before, the equality in (24) captures a tangency between the iso-welfare curves of the home country and foreign country $*1$ in a graph with τ on the y axis and τ^{*1} on the x axis, which we now illustrate in Figure 2. The inequality indicates that the iso-world-price curve, which gives combinations of τ and τ^{*1} that preserve $\widetilde{p}^w(\tau, \tau^{*1}, \tau^{*2})$ at the value obtained at the Horn-Wolinsky MFN solution, is now steeper, as Figure 2 depicts. This ensures that, at the prevailing world price, the home country desires less trade whereas foreign country $*1$ prefers more trade. Finally, since foreign country $*2$ is impacted by changes in τ and τ^{*1} only insofar as those changes impact the world price, we may equivalently express the slope of the iso-world-price curve in terms of the slope of iso-welfare curve for foreign country $*2$.

Starting at a tariff vector that satisfies (15) and (24), and under the assumptions (6) and (16), we now consider the following perturbation:

$$d\tau^1 = \epsilon \tag{25}$$

$$d\tau^2 = \epsilon + \sigma \tag{26}$$

$$d\tau^{*1} = (-\frac{dW}{d\tau}/\frac{\partial W}{\partial \tau^{*1}})\epsilon = (-\frac{dW^{*1}}{d\tau}/\frac{\partial W^{*1}}{\partial \tau^{*1}})\epsilon \tag{27}$$

$$d\tau^{*2} = (-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma \tag{28}$$

where the equality in (27) follows from the bilateral efficiency condition (15) which the starting tariffs are assumed to satisfy, and where $\epsilon > 0$ and $\sigma > 0$ are both small. We give a further condition below concerning the relative magnitudes of ϵ and σ.

Just as before, we can now compute the welfare differentials. For the home country, we get

$$dW = \frac{dW}{d\tau}\epsilon + \frac{\partial W}{\partial \tau^2}\sigma + \frac{\partial W}{\partial \tau^{*1}}(-\frac{dW}{d\tau}/\frac{\partial W}{\partial \tau^{*1}})\epsilon + \frac{\partial W}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma.$$

Thus,

$$dW = [\frac{\partial W}{\partial \tau^2} + \frac{\partial W}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})]\sigma > 0,$$

where the inequality follows since $\sigma > 0$ and (6) holds at the original tariff vector.

For foreign country $*1$, we get

$$dW^{*1} = \frac{dW^{*1}}{d\tau}\epsilon + \frac{\partial W^{*1}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau}/\frac{\partial W^{*1}}{\partial \tau^{*1}})\epsilon + \frac{\partial W^{*1}}{\partial \tau^2}\sigma + \frac{\partial W^{*1}}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma = 0,$$

which confirms that the perturbation maintains the welfare of foreign country $*1$.

Finally, for foreign country $*2$, we get

$$\begin{aligned} dW^{*2} &= \frac{dW^{*2}}{d\tau}\epsilon + \frac{\partial W^{*2}}{\partial \tau^2}\sigma + \frac{\partial W^{*2}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau}/\frac{\partial W^{*1}}{\partial \tau^{*1}})\epsilon + \frac{\partial W^{*2}}{\partial \tau^{*2}}(-\frac{\partial W^{*1}}{\partial \tau^2}/\frac{\partial W^{*1}}{\partial \tau^{*2}})\sigma \\ &= [\frac{dW^{*2}}{d\tau} + \frac{\partial W^{*2}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau}/\frac{\partial W^{*1}}{\partial \tau^{*1}})]\epsilon + [\frac{\partial W^{*2}}{\partial \tau^2} - \frac{\partial W^{*1}}{\partial \tau^2}]\sigma. \end{aligned}$$

For the final expression presented here, we know from (6) that the bracketed expression in the second term is negative. Our next task is to use (6), (16) and (24) to sign the bracketed expression in the first term.

To this end, we recall the inequality in (24):

$$-\frac{\partial W^{*1}}{\partial \tau^{*1}}/\frac{dW^{*1}}{d\tau} < -\frac{\partial W^{*2}}{\partial \tau^{*1}}/\frac{dW^{*2}}{d\tau}.$$

Using (6) and (16), we can re-write this inequality as

$$\frac{dW^{*2}}{d\tau} + \frac{\partial W^{*2}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau}/\frac{\partial W^{*1}}{\partial \tau^{*1}}) > 0,$$

indicating that the bracketed expression in the first term to which we refer above is positive.

It now follows that $dW^{*2} > 0$ if and only if

$$\epsilon > \{\frac{-[\frac{\partial W^{*2}}{\partial \tau^2} - \frac{\partial W^{*1}}{\partial \tau^2}]}{\frac{dW^{*2}}{d\tau} + \frac{\partial W^{*2}}{\partial \tau^{*1}}(-\frac{dW^{*1}}{d\tau}/\frac{\partial W^{*1}}{\partial \tau^{*1}})}\}\sigma. \tag{29}$$

Thus, since the expression in braces is positive, $dW^{*2} > 0$ holds if the second-step perturbation parameterized by σ is small relative to the first-step perturbation parameterized by ϵ.

Consider now the direction of the tariff changes. Given (6) and (16), it is straightforward to see from (25)-(28) that all tariffs are increased except for τ^{*2} which is reduced.

A final point is that we can ensure welfare gains for all countries, including foreign country $*1$, if we augment the perturbation to include an arbitrarily small increase in τ^{*1}. This augmented perturbation raises all three welfare levels.

We now have:

Proposition 7 *For a given bargaining parameter α, suppose the Horn-Wolinsky MFN solution delivers an interior solution, $\widetilde{\boldsymbol{\tau}}_{hw} \equiv (\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw}, \widetilde{\tau}_{hw}, \overline{\tau}^{*2})$, as captured by (15). Suppose at this tariff vector that (6) and (16) hold. Consider case 2 as defined by (24). For sufficiently small $\epsilon > 0$, $\sigma > 0$ satisfying (29), the perturbation defined in (25)-(28) when augmented with an arbitrarily small increase in τ^{*1} has the following effects: all tariffs are raised except for τ^{*2} which is reduced, and all countries enjoy welfare gains.*

Case 3 A final case that might be considered occurs when, at the Horn-Wolinsky MFN solution and the induced MFN world price $\widetilde{p}^w(\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw}, \overline{\tau}^{*2})$, the home country and foreign country $*1$ each achieve their preferred levels of trade.

Formally, this case arises when

$$-\frac{\partial W}{\partial \tau^{*1}} / \frac{dW}{d\tau} = -\frac{\partial W^{*1}}{\partial \tau^{*1}} / \frac{dW^{*1}}{d\tau} = -\frac{\partial W^{*2}}{\partial \tau^{*1}} / \frac{dW^{*2}}{d\tau}, \tag{30}$$

where the expressions are evaluated at the Horn-Wolinsky MFN solution and where total derivatives again reflect the fact that $\tau \equiv \tau^1 = \tau^2$ in the MFN setting. In the case captured by (30), for a graph with τ on the y axis and τ^{*1} on the x axis, the home country iso-welfare curve, the iso-welfare curve of foreign country $*1$ and the iso-world-price locus are all tangent at the tariff pair corresponding to the Horn-Wolinsky MFN solution, as depicted in Figure 3.

One particular example of this kind occurs when the tariff vector corresponds to the MFN politically optimal tariff vector (so that τ^{*2} is set at a particular value, too). Bagwell and Staiger (1999) show that this tariff vector is efficient within the full set of discriminatory tariffs. Thus, as a general matter, it is not always possible in this case to engineer a perturbation that increases welfare for each of the three countries.

We conclude this section with a brief discussion concerning the necessary features of perturbations defined in the full tariff space that generate Pareto gains when we start at the Horn-Wolinsky MFN solution. Recall that in Section 5 we provided such characterizations for a setting in which the MFN requirement is not imposed on the Horn-Wolinsky solution. Unfortunately, once the MFN requirement is imposed, necessary features of Pareto-improving perturbations are more challenging to characterize. Whether or not the MFN requirement is imposed on the solution concept, we have an unlimited space of tariff perturbations to try and restrict. In the absence of an MFN requirement, however, the home country negotiates with each foreign country, and so the Horn-Wolinsky solution concept generates tangency restrictions for each bilateral relationship. By contrast, the Horn-Wolinsky MFN solution doesn't generate a similar tangency restriction between the home-country MFN tariff and the tariff of the non principal supplier country (i.e., foreign country $*2$). We lose some leverage for this reason, making it more difficult to restrict the set of Pareto-improving perturbations.[14]

7 The MFN-Constrained Efficiency Frontier

In the previous section we considered the efficiency properties of the Horn-Wolinsky MFN solution when evaluated relative to the full space of tariff policies $\boldsymbol{\tau} \equiv (\tau^1, \tau^{*1}, \tau^2, \tau^{*2})$. We now assess the

[14] Starting at the Horn-Wolinsky MFN solution and allowing for discriminatory perturbations, suppose that, in a graph with τ^2 and τ^{*2} on the axes, there is an upward lens (not a tangency) formed by the home-country and foreign-country $*2$ iso-welfare curves. Due to the first-order welfare gains that higher (within-lens) values for τ^2 and τ^{*2} generate for the home country and foreign country $*2$, it is then a simple matter to construct a Pareto-improving perturbation in which all four tariffs are increased. Our sufficiency propositions presented in this section, by contrast, are more robust in that they do not impose an assumption about the initial direction of any tariff lens in the relationship between the home country and foreign country $*2$. For general settings, they identify specific perturbations that generate Pareto gains while lowering all tariffs (Proposition 6) or at least some tariff (Proposition 7).

efficiency properties of this solution when judged by the MFN-constrained efficiency frontier. To this end, we ask the following question: Under the assumptions given in (6) and (16), what are the efficiency properties of the Horn-Wolinsky MFN solution $\widetilde{\boldsymbol{\tau}}_{hw} \equiv (\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw}, \widetilde{\tau}_{hw}, \overline{\tau}^{*2})$ as captured by (15) when efficiency is defined relative to the space of MFN tariff policies, $(\tau, \tau^{*1}, \tau, \tau^{*2})$, where $\tau \equiv \tau^1 = \tau^2$ is the home country's MFN tariff?

To answer this question, we now assume that the country welfare functions W and W^{*i} for $i = 1, 2$ satisfy standard regularity conditions so that each point on the MFN-constrained efficiency frontier is interior in that $(\tau, \tau^{*1}, \tau, \tau^{*2}) \in (\underline{\tau}, \overline{\tau})^4$, and we begin by exploiting an implication of the characterization of the MFN-constrained efficiency frontier derived in Bagwell and Staiger (2005). According to their Proposition 7, at an MFN-efficient vector of tariffs, either

$$-\frac{\partial W^{*1}}{\partial \tau^{*1}} / \frac{dW^{*1}}{d\tau} < -\frac{\partial W}{\partial \tau^{*1}} / \frac{dW}{d\tau} < -\frac{\partial \widetilde{p}^w}{\partial \tau^{*1}} / \frac{\partial \widetilde{p}^w}{\partial \tau}; \tag{31}$$

$$-\frac{\partial \widetilde{p}^w}{\partial \tau^{*1}} / \frac{\partial \widetilde{p}^w}{\partial \tau} < -\frac{\partial W}{\partial \tau^{*1}} / \frac{dW}{d\tau} < -\frac{\partial W^{*1}}{\partial \tau^{*1}} / \frac{dW^{*1}}{d\tau}; \text{ or} \tag{32}$$

$$-\frac{\partial \widetilde{p}^w}{\partial \tau^{*1}} / \frac{\partial \widetilde{p}^w}{\partial \tau} = -\frac{\partial W}{\partial \tau^{*1}} / \frac{dW}{d\tau} \gtrless -\frac{\partial W^{*1}}{\partial \tau^{*1}} / \frac{dW^{*1}}{d\tau}, \tag{33}$$

where total derivatives reflect that $\tau \equiv \tau^1 = \tau^2$ in the MFN setting. In writing these expressions, we use the fact that foreign country $*2$ is impacted by changes in τ and τ^{*1} only insofar as those changes impact the world price, so that in (τ, τ^{*1}) space we may equivalently express the slope of the iso-welfare curve for foreign country $*2$ as the slope of the iso-world-price curve.

At points on the MFN-constrained efficiency frontier satisfying (31), the home country desires less trade whereas foreign country $*1$ prefers more trade at the prevailing world price. At points on the MFN-constrained efficiency frontier satisfying (32), the home country desires more trade whereas foreign country $*1$ prefers less trade at the prevailing world price. And finally, at points on the MFN-constrained efficiency frontier satisfying (33), the home country achieves its desired trade volume whereas foreign country $*1$ may desire more trade, desire less trade, or achieve its desired trade volume as well at the prevailing world price. Notice that under condition (31), the home country's iso-welfare curve is steeper than that of foreign country $*1$, indicating a downward lens between the home country and foreign country $*1$ in (τ, τ^{*1}) space, while under condition (32) the home country's iso-welfare curve is flatter than that of foreign country $*1$, indicating an upward lens between the home country and foreign country $*1$. Finally, under condition (33), there can be a downward lens, an upward lens, or no lens at all between the home country and foreign country $*1$, with this last possibility corresponding to the MFN politically optimal tariffs.

Consider now three cases. A first case is where τ^{*2} is fixed at a level $\overline{\tau}^{*2}$ that is consistent with a point on the MFN-constrained efficiency frontier that satisfies (31). In this case, as we have observed, there is a downward lens between the home country and foreign country $*1$. But the Horn-Wolinsky MFN solution is characterized by a point of tangency between the home-country and foreign-country $*1$ iso-welfare curves as described by (15), so it follows that for this case there exists a range of values for the initial tariff vector $\boldsymbol{\tau}_0$ and the bargaining parameter α such that the Horn-Wolinsky MFN solution corresponds to a point of tangency within this lens and therefore entails greater liberalization between the home country and foreign country $*1$ than is MFN-constrained efficient given $\tau^{*2} \equiv \overline{\tau}^{*2}$. A second case is where τ^{*2} is fixed at a level $\overline{\tau}^{*2}$ that is consistent with a point on the MFN-constrained efficiency frontier that satisfies (32). In this case, as we have observed, there is an upward lens between the home country and foreign country $*1$. With the Horn-Wolinsky MFN solution described by (15), it then follows that for this case there exists a range of values for the initial tariff vector $\boldsymbol{\tau}_0$ and the bargaining parameter α such that the Horn-Wolinsky MFN solution corresponds to a point of tangency within this lens and therefore entails less liberalization between the home country and foreign country $*1$ than is MFN-constrained efficient given $\tau^{*2} \equiv \overline{\tau}^{*2}$. A third case is where τ^{*2} is fixed at a level $\overline{\tau}^{*2}$ that is consistent with a point on the MFN-constrained efficiency frontier that satisfies (33). As we have observed, in this case there can be a downward lens, an upward lens, or no lens at all between the home country and foreign country $*1$, with this last possibility corresponding to the MFN politically optimal tariffs. It then follows that for this case the Horn-Wolinsky MFN solution may entail less liberalization or more liberalization between the home country and foreign country $*1$ than is MFN-constrained efficient given $\tau^{*2} \equiv \overline{\tau}^{*2}$; or the Horn-Wolinsky MFN solution could achieve the MFN-efficient political optimum.

This last possibility is of some special interest. Evidently, if τ^{*2} is fixed at its MFN politically optimal level, there exists some initial tariff vector $\boldsymbol{\tau}_0$ and some bargaining parameter α such that the Horn-Wolinsky MFN solution for the bargain between the home country and foreign country $*1$ would deliver countries to the MFN-efficient political optimum. Of course, this is an extremely special set of conditions. More generally, we can conclude that, with the exception of this knife-edge case, the Horn-Wolinsky MFN solution will be inefficient, and it can lead to either too much liberalization or too little liberalization relative to the MFN-constrained efficiency frontier.

We summarize with:

Proposition 8 *For a given bargaining parameter α, suppose the Horn-Wolinsky MFN solution delivers an interior solution, $\widetilde{\boldsymbol{\tau}}_{hw} \equiv (\widetilde{\tau}_{hw}, \widetilde{\tau}^{*1}_{hw}, \widetilde{\tau}_{hw}, \overline{\tau}^{*2})$, as captured by (15). Suppose at this tariff vector that (6) and (16) hold. And finally, suppose that each point on the MFN-constrained efficiency frontier is interior. Then the Horn-Wolinsky MFN solution is generically inefficient relative to the MFN-constrained efficiency frontier, and may lead to either too little liberalization or too much liberalization relative to MFN-constrained efficient levels.*

In light of the positive externality that the home country's MFN tariff liberalization imparts on foreign country $*2$, it may seem surprising that the Horn-Wolinsky MFN solution could ever lead to too much liberalization relative to MFN-constrained efficient tariff levels. But recalling that, beginning from MFN-efficient tariff levels, the home country and foreign country $*1$ can only gain in their bilateral bargain if they worsen foreign country $*2$'s – and hence under MFN, also foreign country $*1$'s – terms of trade, the bilateral bargain must move foreign country $*1$'s trade volume in the direction that foreign country $*1$ would desire at a fixed terms of trade; and if at the relevant point on the MFN-constrained efficiency frontier foreign country $*1$ desires more trade volume at a fixed terms of trade, then according to the Horn-Wolinsky MFN solution the bilateral bargain between the home country and foreign country $*1$ will lead these countries to reduce their tariffs and engage in too much liberalization relative to MFN-constrained efficient tariff levels.[15]

8 Discussion

In the preceding sections we have applied the Horn-Wolinsky bargaining solution to a setting of simultaneous bilateral tariff bargaining, both when tariffs can be discriminatory and when they must conform to the MFN rule, and we have evaluated the efficiency properties of this solution. In the context of the GATT/WTO, the properties of the Horn-Wolinsky MFN solution we have characterized are especially relevant. Here we describe two additional features of GATT/WTO tariff bargaining from which our analysis above has abstracted, and we discuss the potential importance of and also the challenges in extending the Horn-Wolinsky solution to environments that incorporate these features.

MFN bargaining beyond the principal supplier In our analysis of the efficiency properties of the Horn-Wolinsky solution in the presence of MFN, where the home country must apply the same tariff across its trading partners, we have assumed that the home country only bargains with its principal supplier. Analyzing the detailed United States bargaining records of the 1950-51 GATT Torquay Round, Bagwell, Staiger and Yurukoglu (2016) confirm that engaging a single exporting country in negotiations over an MFN import tariff is the modal US behavior. However, Bagwell, Staiger and Yurukoglu also report that there are significant numbers of US bargains on a given tariff that involve more than one exporting country. Here we briefly comment on the challenges involved in extending the Horn-Wolinksy solution concept to encompass these broader bargaining possibilities.

In principle, the Horn-Wolinsky solution concept could be applied under MFN in the case where the home country bargains with both of its trading partners by introducing to the model the notion of MFN tariff *bindings* – legal commitments to a maximum applied tariff level – and allowing the home country to negotiate different binding levels on its import tariff with its different bargaining partners, with the understanding that its applied tariff cannot exceed the minimum binding level over the set of bindings it agrees to in its bargains. In this environment, an issue of multiplicity of Horn-Wolinsky equilibria arises, however. To see this, let us denote the binding on its tariff

[15] An interesting question is whether more can be said about the conditions under which each of the possibilities described in Proposition 8 might arise. While we can't offer a complete answer to this question here, if we think of the tariff of foreign country $*2$ as being fixed on its reaction curve, and if governments are sufficiently close to national income maximizers, then it is straightforward to establish that, at the relevant point on the MFN-constrained efficiency frontier, foreign country $*1$ desires more trade volume at the fixed terms of trade while the home country desires less trade volume at the fixed terms of trade; and hence, under these conditions, the relevant case is indeed likely to correspond to the case of too much liberalization relative to the MFN-constrained efficient tariff levels.

τ that home negotiates with $*1$ by τ_{b1} and the binding on its tariff τ that home negotiates with $*2$ by τ_{b2}, with the understanding that home's tariff will then be applied at $\tau = \min[\tau_{b1}, \tau_{b2}]$ under the assumption that these binding levels are positioned below home's best-response tariff. The key point is that the relevant lenses will be defined over the applied tariffs rather than the bindings, and that any positioning of the bindings that would result in an upward-directional lens for each bilateral will be a Horn-Wolinsky equilibrium, because the binding in one bilateral will keep home from moving into the lens in the other bilateral, and vice versa. To usefully apply the Horn-Wolinsky solution concept under MFN while permitting multiple bargaining partners for the home country, this multiplicity of equilibria will have to be addressed in some way, an issue that we leave for future research.

Article XXVIII renegotiation In our analysis of the efficiency properties of the Horn-Wolinsky solution in the context of bilateral simultaneous tariff bargaining, we have abstracted from the renegotiation possibilities that are provided under GATT/WTO rules. But Bagwell and Staiger (1999, 2002) have argued that the particular renegotiation provisions included in GATT Article XXVIII have the effect of ensuring that no country can be forced in a bilateral GATT/WTO bargain to accept greater trade volume than it desires at the given terms of trade/world price. This is an interesting feature of the GATT/WTO bargaining setting to consider in this context, because intuitively the possibility of Article XXVIII renegotiation could diminish the amount of negotiated trade liberalization that bargaining partners can achieve. In light of our results above concerning the Horn-Wolinsky solution in the absence of MFN and in its presence, this suggests that the impact of the possibility of Article XXVIII renegotiation might diminish the amount of excessive liberalization in the absence of MFN and thereby move the outcome toward the efficiency frontier, while it could either move the outcome toward or away from the efficiency frontier in the presence of MFN.[16]

While the intuition for the above statements seems strong, the method of analysis to confirm this intuition would have to be quite different from that which we have pursued above. In the preceding sections, we have relied heavily on the fact that at a Horn-Wolinsky solution each bilateral satisfies a tangency condition for the indifference curves of the two bargaining parties when the policies determined in the other bilateral are taken as fixed. With Article XXVIII renegotiation, however, this tangency condition will in general not be met. A different approach to constructing perturbations from the Horn-Wolinsky solution is therefore necessary to proceed with an analysis that includes the possibility of Article XXVIII renegotiation. We leave this important task for future research.

[16] Renegotiation under GATT Article XXVIII is conducted with respect to MFN tariff bindings, but it is also interesting to consider the implications of bilateral renegotiation in a setting with discriminatory tariffs.

9 Conclusion

We consider a three-country, two-good model of bilateral tariff negotiations where each country is affected by the outcomes achieved in each bilateral negotiation. We characterize the negotiated tariffs that are predicted by the Horn-Wolinsky solution, both in an unrestricted setting where discriminatory tariffs are allowed and in a restricted setting where tariffs must satisfy the MFN rule. Our main objective is to characterize the efficiency properties of the negotiated tariffs. For the unrestricted setting, we show that starting from an interior Horn-Wolinsky solution we can construct a Pareto-improving perturbation under which all tariffs are increased. We also characterize the necessary features of Pareto-improving perturbations. For the restricted (MFN) setting, we show for one important case that, starting from a Horn-Wolinsky MFN solution, we can construct a Pareto-improving perturbation under which all tariffs are reduced. We also provide characterizations of the efficiency of the Horn-Wolinsky MFN solution relative to the MFN-constrained efficiency frontier.

The Nash-in-Nash approach of the Horn-Wolinsky solution underlies a large and important body of applied work in Industrial Organization that studies surplus division in bilateral oligopoly settings. Our work here provides a theoretical foundation for related studies in International Trade that address bilateral tariff negotiations. In addition to such applied work, our work motivates further examination of the micro-foundation of the Nash-in-Nash solution for settings in which negotiated outcomes go beyond surplus division and impact efficiency.

10 References

Bagwell, Kyle and Robert W. Staiger (1999), "An Economic Theory of GATT," *American Economic Review* 89(1): 215-48.

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

Bagwell, Kyle and Robert W. Staiger (2005), "Multilateral Trade Negotiations, Bilateral Opportunism and the Rules of GATT/WTO," *Journal of International Economics* 67(2): 268-94.

Bagwell, Kyle and Robert W. Staiger (2010), "Backward Stealing and Forward Manipulation in the WTO," *Journal of International Economics* 82(1): 49-62.

Bagwell, Kyle and Robert W. Staiger (2016), "Multilateral Trade Bargaining and Dominant Strategies," NBER Working Paper 22842.

Bagwell, Kyle, Robert W. Staiger and Ali Yurukoglu (2016), "Multilateral Trade Bargaining: A First Look at the GATT Bargaining Records," working paper.

Chan, Jackie M. L. (2015), "Strategic Tariff Bargaining in the GATT and Gradualism," working paper.

Chipty, Tasneem and Chris M. Snyder (1999), “The Role of Firm Size in Bilateral Bargaining: A Study of the Cable Television Industry,” *Review of Economics and Statistics*, 81(2), 326-40.

Collard-Wexler, Allan, Gautam Gowrisankaran, and Robin S. Lee (2016), “‘Nash-in-Nash Bargaining: A Microfoundation for Applied Work,” working paper.

Crawford, Greg, Robin S. Lee, Michael D. Whinston and Ali Yurukoglu (2015), “The Welfare Effects of Vertical Integration in Multichannel Television Markets,” working paper.

Crawford, Greg and Ali Yurukoglu (2012), “The Welfare Effects of Bundling in Multi-channel Television Markets,” *American Economic Review*, 102(2), 643-85.

Cremer, J. and M. H. Riordan (1987), “On Governing Multilateral Transactions with Bilateral Contracts,” *RAND Journal of Economics*, 18, 436-51.

Dasgupta, Partha and Eric Maskin (2015), “Debreu’s Social Equilibrium Existence Theorem,” *Proceedings of the National Academy of Sciences* 112 (52):15769-15770.

Debreu Gerard (1952), “A Social Equilibrium Existence Theorem,” *Proceedings of the National Academy of Sciences* 38(10):886–893.

Friedman, James (1990), *Game Theory with Applications to Economics*, Oxford University Press.

Gowrisankaran, Gautum, Aviv Nevo and Robert Town (2015), “Mergers when Prices Are Negotiated: Evidence from the Hospital Industry,” *American Economic Review*, 105(1), 172-203.

Grennan, Matthew (2013), “Price Discrimination and Bargaining: Empirical Evidence from Medical Devices,” *American Economic Review*, 103(1), 147-77.

Hart, Oliver and Jean Tirole (1990), “Vertical Integration and Market Foreclosure,” *Brookings Papers on Economic Activity, Microeconomics*, 205-76.

Ho, Kate and Robin S. Lee (2015), “Insurer Competition in Health Care Markets,” NBER Working Paper 19401.

Horn, Henrik and Asher Wolinsky (1988), “Bilateral Monopolies and Incentives for Merger,” *RAND Journal of Economics* 19(3): 408-419.

McAfee, R. Preston and Marius Schwartz (1994), “Opportunism in Multilateral Vertical Contracting: Nondiscrimination, Exclusivity, and Uniformity,” *American Economic Review*, 84, 210-30.

McAfee, R. Preston and Marius Schwartz (1995), “The Non-existence of Pairwise-proof Equilibria,” *Economics Letters*, 49, 251-59.

Nash John F (1950), “Equilibrium Points in N-person Games,” *Proceedings of the National Academy of Sciences* 36(1):48–49.

Vives, Xavier (1999), *Oligopoly Pricing. Old Ideas and New Tools*, Cambridge, MA: The MIT Press.

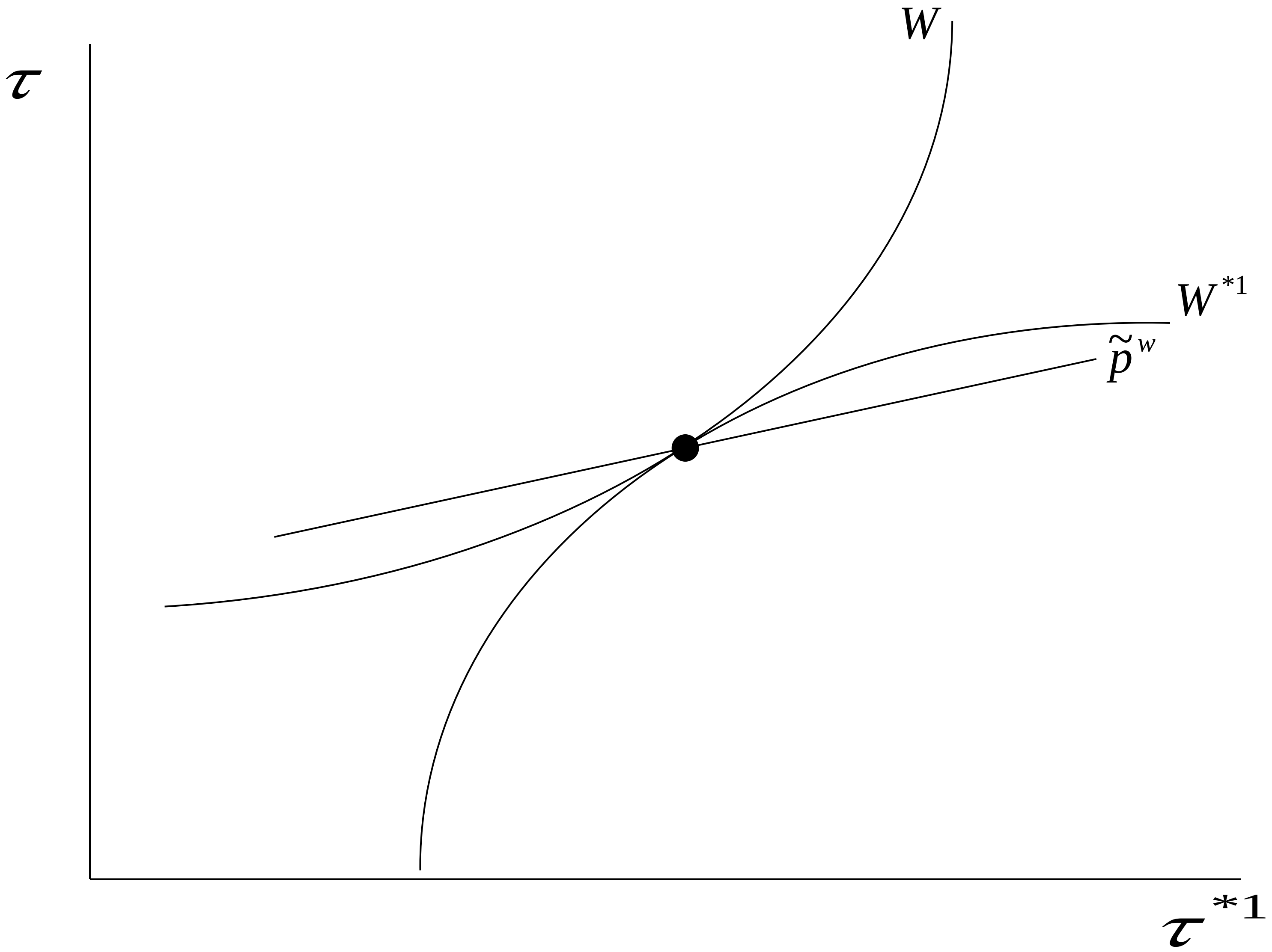

Figure 1

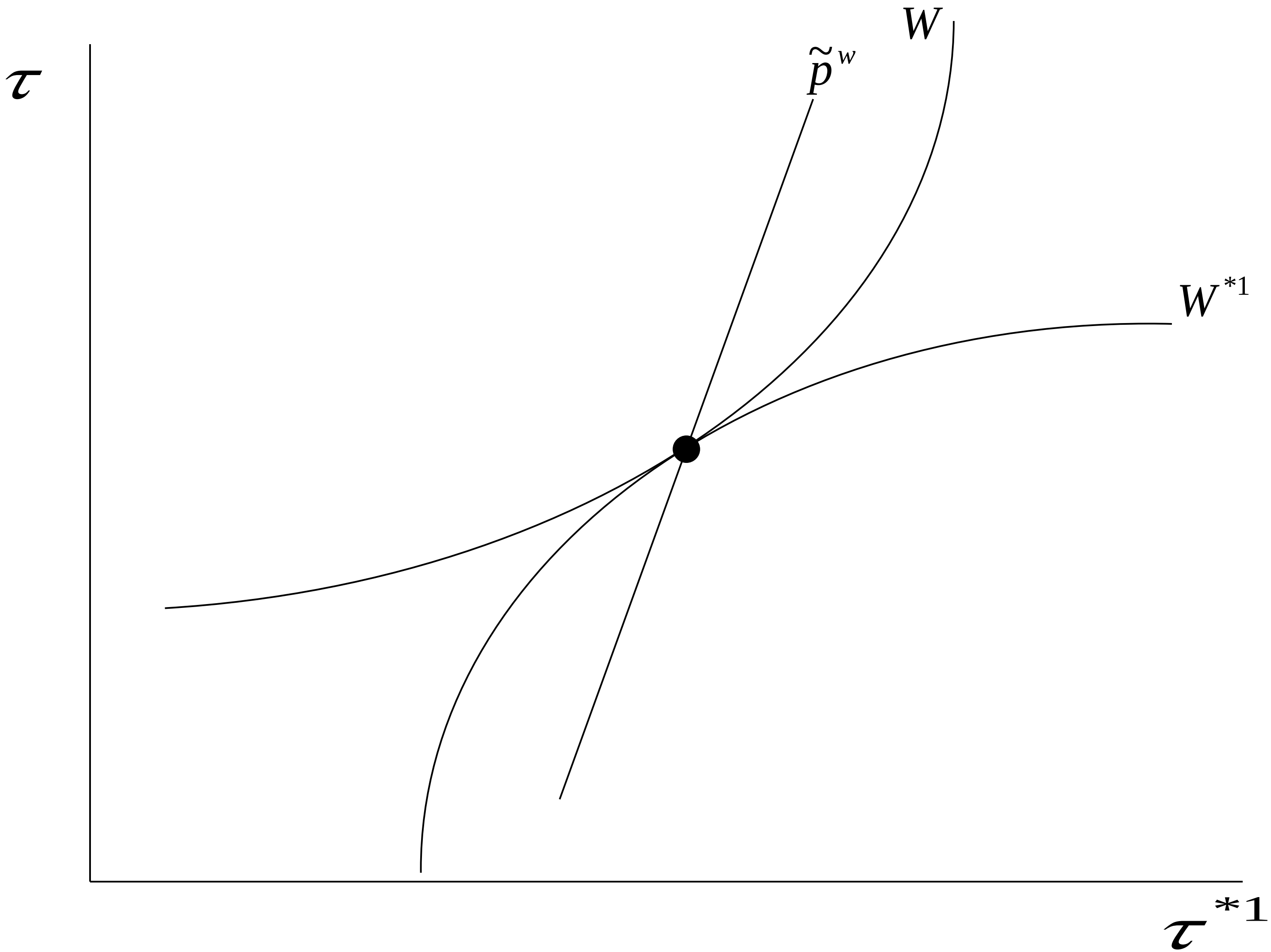

Figure 2

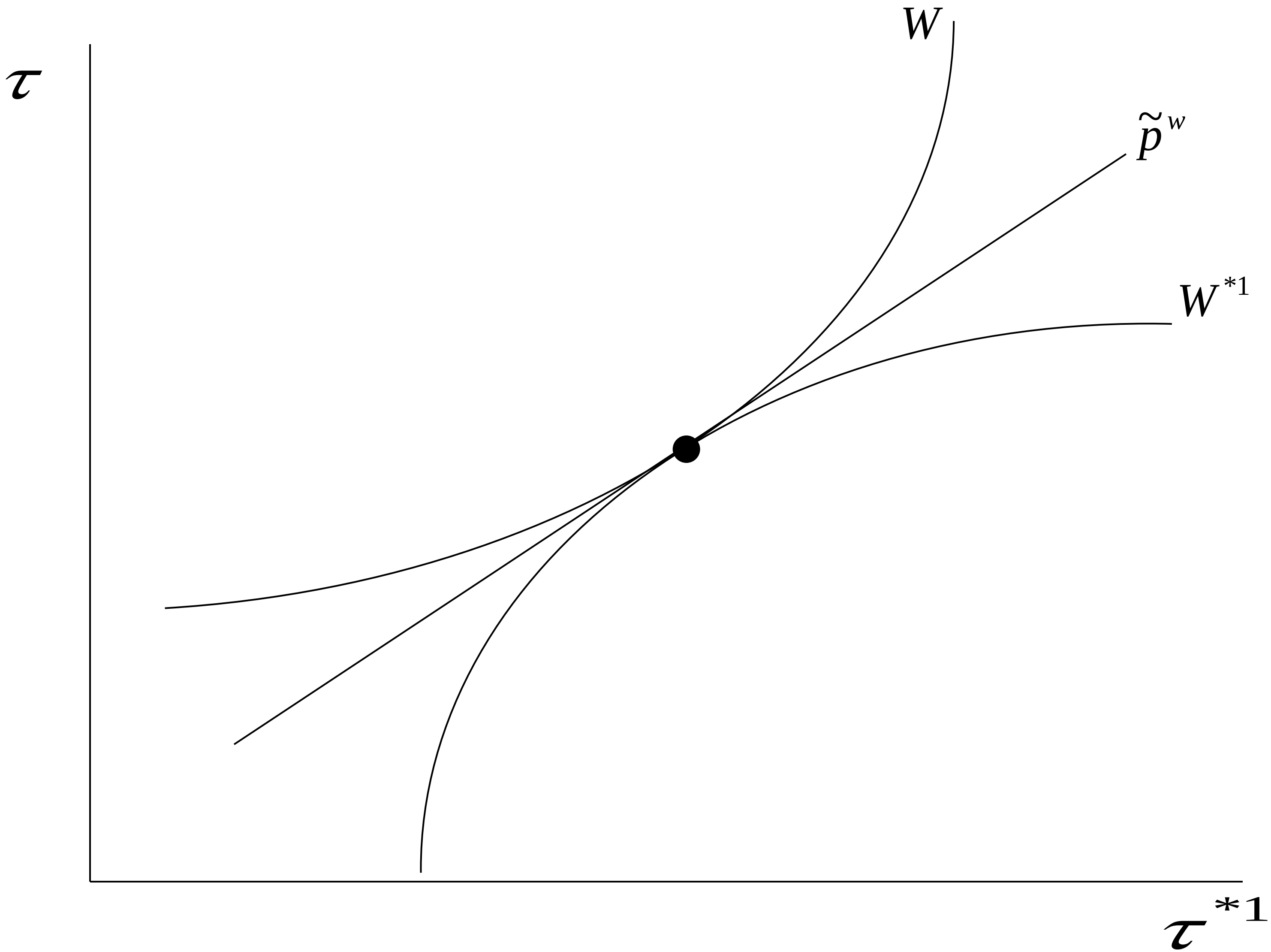

Figure 3