“Nash-in-Nash” tariff bargaining

Kyle Bagwell \(^{a,d,a}\), Robert W. Staiger \(^{b,d}\), Ali Yurukoglu \(^{c,d}\)

\(^{a}\) Department of Economics, Stanford University, USA
\(^{b}\) Department of Economics, Dartmouth College, USA
\(^{c}\) Graduate School of Business, Stanford University, USA
\(^{d}\) NBER, USA

FULL LENGTH ARTICLES

ABSTRACT

We provide an equilibrium analysis of the efficiency properties of simultaneous bilateral tariff negotiations in a three-country model of international trade. We consider the setting in which discriminatory tariffs are allowed, and we utilize the “Nash-in-Nash” solution concept of Horn and Wolinsky (1988). We allow for a general family of political-economic country welfare functions and assess efficiency relative to these welfare functions. 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.

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. Furthermore, the outcome of any one bilateral trade negotiation naturally may be impacted by the anticipated outcomes in other contemporary bilateral negotiations.

In this paper, we provide an equilibrium analysis of the efficiency properties of simultaneous bilateral tariff negotiations in a three-country model of international trade. Specifically, the trade model entails a home country and two foreign countries, where the home country trades with both foreign countries but the foreign countries do not trade with one another. To analyze simultaneous bilateral tariff negotiations in this context, two basic questions must be answered. First, what, if any, rules are imposed on the negotiated bilateral tariffs? Second, what is the equilibrium concept that governs the manner in which simultaneous bilateral tariff agreements are negotiated?

Negotiation rules could take many forms, and the implications of different rules for trade-bargaining outcomes is an important subject. Our focus here, however, is different. Rather than study the potential implications of different rules, we seek to provide an equilibrium analysis of the “no-rules” benchmark case in which discriminatory tariffs of any non-prohibitive value are allowed. Thus, in our analysis, the home country is free to negotiate bilateral agreements under which it applies discriminatory tariffs to imports from its two foreign trading partners.

We may motivate consideration of the no-rules setting in two ways. First, tariff discrimination is an important possibility worthy of study in its own right, as it is a feature of various important historical trading relationships and arises to varying degrees in the current era as well among GATT/WTO member countries.\(^1\) Second, by offering an equilibrium analysis of the non-rules setting, we provide a valuable benchmark for future work that may extend the equilibrium analysis developed here to include rules such as the most-favored nation (MFN) rule.\(^2\)

We turn now to the equilibrium concept. A basic feature of the no-rules setting is that the negotiated tariffs in any one bilateral relationship may affect world prices and thus the payoffs (i.e., welfare levels) that are associated with negotiated tariffs in the other bilateral

---

\(^1\) See Pomfret (1997, Part 1) for a detailed history of discriminatory trade policies in the 19th and 20th centuries up to the early years of the WTO. Pomfret (1997, p. 30) and Viner (1951, p. 259) argue that the general trend of trade barriers over the period from the late 1870s until the Second World War was marked by an increase in their discriminatory character. As Beckett (1941, pp. 24-30), Pomfret (1997, p. 22), Rohlfing (2009), Tasca (1938, pp. 146-7) and Tavares (2006) discuss, tariff discrimination may occur even among trading relationships governed by the most-favored nation (MFN) rule, if countries limit the breadth of the MFN rule and impose narrow product reclassifications that can facilitate the application of different tariffs on imports of broadly similar goods from different trading partners. Beckett (1941, p.28), for example, reports that approximately 1/3 of the tariff paragraphs modified by concessions negotiated by the United States under the Reciprocal Trade Agreement Act were associated with product reclassifications. Discrimination is also explicitly allowed under certain GATT/WTO rules to allow for preferential trading agreements, preferences for developing countries and plurilateral agreements in which a subset of countries extend certain benefits to one another.

\(^2\) In the Conclusion, we discuss this extension and some of the modeling challenges that it raises.

https://doi.org/10.1016/j.jinteco.2019.103263
0022-1996/© 2019 Elsevier B.V. All rights reserved.
relationship. The bilateral negotiations are then fundamentally interdependent. We thus require an equilibrium model of simultaneous bilateral tariff negotiations when payoffs and hence negotiation outcomes are interdependent. Our approach here is to adopt the equilibrium solution concept of Horn and Wolinsky (1988).

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.3 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.4 To our knowledge, we are the first to explore the theoretical implications of the Nash-in-Nash approach for simultaneous bilateral tariff negotiations.

The Nash-in-Nash approach offers two main benefits. First, using this approach, we can capture in an equilibrium framework some intuitive strategic features of simultaneous bilateral negotiations with interdependent payoffs. We note in particular that, under the Nash-in-Nash approach, each negotiating pair fails to internalize the consequences of its bargaining outcome for the other negotiating pair. Second, the Nash-in-Nash approach offers a tractable means of characterizing bargaining outcomes when bilateral negotiations are simultaneous and payoffs are interdependent. Correspondingly, and 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.

At the same time, a limitation of the Nash-in-Nash approach 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 within-negotiation communication between agents associated with the same player may be feasible.5,6 The interpretation is arguably less strained, however, in settings with bargaining frictions such that opportunities for communication and coordination arise only after bilateral bargaining positions have hardened.

On balance, we believe that the 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 settings. Focusing on the no-rules setting, our paper provides 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 but are unable to communicate with each other during the course of their respective bilateral negotiations.

To conduct our formal analysis, we use a trade model with a home country and two foreign countries. Our modeling framework includes the scenario of a neoclassical model with two goods as analyzed by Bagwell and Staiger (2005, 2010, 2018) in which the home country imports the same good from each foreign country and exports a second good to the foreign countries. But our assumptions are sufficiently general to include other scenarios as well, such as when the home country imports substitute goods from the foreign countries and perhaps likewise exports substitute goods to the respective foreign countries. Since the foreign countries may export the same or close substitute goods, the model allows for the possibility of discriminatory tariffs for the home country. We thus represent country welfare functions in reduced form as general functions of the four tariffs (two tariffs for the home country, and one tariff for each foreign country) imposed by the three countries. We assess efficiency relative to these country welfare functions, where a country’s welfare function may include both economic and distributional concerns.7

We begin our formal analysis by defining an interior Horn–Wolinsky solution for our no-rules setting. We then assume the existence of an interior Horn–Wolinsky solution and characterize its efficiency properties. 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 are increased 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 an interior Horn–Wolinsky solution. 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. We do show, however, that, if all countries enjoy weak welfare gains under a small 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 harm the welfare of the non-participating country. Using this finding, we then show that, if under a small perturbation all countries enjoy

---

3 For example, Crawford and Yurukoglu (2012) and Crawford et al. (2018) explore negotiations between cable television distributors and content creators, while Brennan (2013), Gowrisankaran et al. (2015), and Ho and Lee (2017) consider negotiations between hospitals, medical device manufacturers, and health insurers.

4 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; Hart and Tirole, 1990). See McAfee and Schwartz (1995) for further discussion.

5 Collard-Wexler et al. (2019) 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.

6 In their study of the GATT Torquay Round, Bagwell et al. (forthcoming) highlight the impact of failed bargains between the United States and several British Commonwealth countries 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 (2018) 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. The Nash-in-Nash approach seems better suited for simultaneous bilateral negotiations under MFN within multilateral GATT/WTO rounds when multilateral reciprocity is not required, as Bagwell et al. (2018) argue.

7 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.”
weak welfare gains and at least one country strictly gains, 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.

Hence, starting at an interior Horn–Wolinsky solution, it is not possible to make every country better off with a small perturbation that induces a weak reduction in the tariffs of every country, but it is possible to make every country better off with a perturbation that generates a small increase in the tariffs of every country. Based on these findings, we conclude that simultaneous bilateral tariff negotiations are associated with excessive liberalization when judged relative to the preferences of all countries. We are not aware of a previous equilibrium analysis that establishes this conclusion.

Turning now to the related literature, it is interesting to compare our results to those in a large literature that examines the possible third-party effects of preferential trading agreements. This literature imposes a significant restriction on the family of discriminatory tariffs (so that trade is free among preferred partners) and then explores different questions such as whether such agreements facilitate or hinder the achievement of global free trade.8 We include as a special case the possibility that countries maximize national income, and for this special case global free trade is of course efficient. Our finding of excess liberalization even for national-income-maximizing countries arises because we allow countries to pursue bilateral agreements in which they exchange discriminatory import subsidies and liberalize beyond free trade.9

The 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 both sells to and buys from its trading partner). Our focus on efficiency is also novel and is appropriate given our aim to study the welfare implications of bilateral tariff negotiations.

A quantitative analysis related to a number of the themes we explore here is contained in Bagwell et al. (2018a). In that paper, we embed a multi-sector model of trade between multiple countries into a model of inter-connected bilateral negotiations over tariffs, where the tariff negotiations are modeled according to the Nash-in-Nash approach. There we quantify the third-party externalities that are central to the theoretical findings described above, and we show that the distinct nature of these externalities with and without MFN is key to understanding the efficiency properties of the Horn–Wolinsky solutions under the different bargaining protocols that we report in that paper.

We note also that an extensive trade-policy literature uses Nash tariffs as a benchmark for understanding the value of multilateral tariff cooperation in GATT and now the WTO. The Nash benchmark is appealing, since it offers a coherent representation of the outcome that would be expected in the absence of any cooperation. At the same time, and as is well known, bilateral tariff agreements were also a feature of the pre-GATT era. From this perspective, an alternative theoretical benchmark in which countries simultaneously pursue bilateral tariff agreements is also appealing. Our analysis offers such an analysis for the no-rules setting with discriminatory tariffs. Valuable future work may explore how the theoretical results vary when the MFN rule is added.10

Finally, our work is closely related to a sequence of papers by Bagwell and Staiger (2005, 2010, 2018). These papers use a neoclassical trade model with three countries and two goods. As mentioned above, this model is included in the modeling framework of the present paper. The papers then go on to develop different insights about bilateral tariff negotiations in this modeling context.

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 MFN rule can be interpreted in this context. As Bagwell and Staiger (2005) argue, however, the MFN rule 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.11

We note that 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. Thus, unlike the present paper, they do not offer an equilibrium analysis of bargaining outcomes.

Bagwell and Staiger (2010) consider rules under which efficient outcomes can be achieved in a subgame perfect equilibrium of 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.12 The present paper differs in several respects: we analyze simultaneous bilateral tariff negotiations, do not allow transfers, consider a no-rules setting, use the Horn–Wolinsky solution concept and show that efficiency is then infeasible since equilibrium tariffs are too low. Finally, Bagwell and Staiger (2018) characterize the outcomes that can be achieved in a multilateral bargaining setting in which proposals are simultaneously made and any proposed outcome must satisfy the MFN rule along with the principle of multilateral reciprocity. As they show, in this “strong-rule” setting, countries are unable to alter the terms of trade, and as a consequence multilateral bargaining outcomes may be characterized while requiring only that countries make dominant-strategy proposals. They show that an efficient outcome can be achieved if and only if the initial tariff vector is such that the world price takes a particular value; otherwise, the resulting tariffs are higher than efficient.

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 definition of an interior Horn–Wolinsky solution. Section 4

---

8 See, for example, Saggi and Yildiz (2010) and Saggi et al. (2013) for analyses of the endogenous formation of preferential trading agreements.

9 While import subsidies are not commonly observed, the Lerner symmetry theorem ensures that the effects of an import subsidy can be equivalently generated by an export subsidy. Pomfret (1997, p. 56) mentions an interesting example in this context, noting that Italy, Austria and Hungary reached an agreement in 1931–2 to subsidize their exports to each other. We also note that, in the context of free trade agreements that set tariffs to zero on substantially all trade, additional “deep integration” commitments in some cases might play a role similar to the role of import subsidies in our formal model.

10 Historically, the MFN rule has taken unconditional and conditional forms. The unconditional MFN rule is enshrined in GATT Article I and is the form typically studied by trade economists. When two countries reach an agreement with a conditional MFN clause, each agrees to extend to the other any tariff cuts established for a third country, conditional on receiving compensation that is equivalent to that provided by the third country. Trade agreements between European countries have featured the unconditional MFN rule since 1860, although France adopted the conditional form for a period that includes the early 1920s. The US adopted the conditional MFN rule until 1923, at which time it embraced the unconditional MFN rule. As Pomfret (1997, pp. 18-19) explains, the US conditional MFN policy was applied in an arbitrary way, so that in practice most-favored treatment was not ensured at all. See also Hawkins (1951, pp. 67-8). When we refer to the MFN rule elsewhere in the paper, we refer to the standard unconditional form of the rule.

11 Bagwell and Staiger (2005) 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, 2002, 2005, 2018) and Ossa (2014).

12 See also Chan (2019).
contains our construction of a Pareto-improving perturbation relative to an interior Horn–Wolinsky solution, and Section 5 provides related findings concerning the necessary features of Pareto-improving perturbations. Section 6 concludes. The Appendix gathers proofs not contained in the body of the paper.

2. Trade model

In this section, we describe the structure of our trade model. In this general context, we then briefly describe the related analysis of Bagwell and Staiger (2005). Finally, we highlight the distinct goal and approach of our analysis.

2.1. The three-country trade model

We consider a three-country model of trade. The model features one home country and two foreign countries, where each foreign country trades only with the home country. As usual, foreign country variables are denoted with an asterisk. In its trading relationship with foreign country \(i\), where \(i = 1, 2\), the home country exports good \(y_i\) and imports good \(x_i\). Following Bagwell and Staiger (2005), our approach allows for the scenario in which the two foreign countries export a homogeneous good, \(x = x_1, x_2\), and the home country likewise exports a homogenous good, \(y = y_1, y_2\). At the same time, our modeling approach is sufficiently general that we can also include the scenario in which the two foreign countries export goods that are imperfect substitutes, \(x_1 \neq x_2\), and the home country likewise exports goods that are imperfect substitutes, \(y_1 \neq y_2\).

The ad valorem import tariff that the home country applies to exports of good \(x_i\) from foreign country \(i\) is denoted as \(t_i\), and the ad valorem import tariff that foreign country \(i\) applies to exports of good \(y_j\) from the home country is denoted as \(t_{ji}\). We define \(t_i', t_{ji}', \tau_i', \tau_{ji}'\), and \(\tau^{*}_i, \tau^{*}_{ji}\) with the resulting tariff vector defined as \(\tau = (\tau_i', \tau_{ji}', \tau_i, \tau_{ji})\). Throughout, we assume that the tariffs are non-prohibitive. Importantly, we allow that \(\tau_i' \neq \tau_{ji}'\) and thus permit the home country to impose different tariffs on the export goods of foreign countries \(1\) and \(2\). We thus interpret our analysis as allowing for discriminatory tariffs. This interpretation is straightforward in the scenario in which the home country imports the same good for discriminatory tariffs. Likewise, in the scenario where the foreign countries export goods that are imperfect substitutes, \(x_1 \neq x_2\), and the home country like-wise exports goods that are imperfect substitutes, \(y_1 \neq y_2\).

We assume that each country has a continuously differentiable welfare function defined over the tariff vector. Formally, letting \(W(\tau)\) represent the welfare of the home country and \(\tau^{*}_i, (\tau')\) represent the welfare of foreign country \(i\) for \(i = 1, 2\), we assume that \(W(\tau)\) and \(\tau^{*}_i, (\tau')\) are continuously differentiable functions. For any given country, the corresponding welfare function could represent the national welfare for that country. Other interpretations are also possible, however. Indeed, as Bagwell and Staiger (1999, 2002) argue in detail for the two-good model (i.e., for the first scenario described above), the welfare function for a given country could also reflect political-economic considerations such as distributional concerns or the influence of lobbies. We present further assumptions on the welfare functions in Section 4.

2.2. Bagwell and Staiger’s (2005) Analysis

Bagwell and Staiger (2005) explore a neoclassical model with three countries and two goods, \(x\) and \(y\), corresponding to the first scenario described above. They represent each government’s welfare as a function of the relative local price in its country and the given country’s terms of trade, where the home country’s terms of trade is given by a multilateral measure. Since the market-clearing local and world prices are ultimately determined by the tariff vector \(\tau\), the welfare functions can be represented in reduced form as functions of \(\tau\), so that the home- and foreign-country welfare functions are given as \(W(\tau)\) and \(W^{*}_i(\tau)\), respectively, where \(i = 1, 2\) and where \(W(\tau)\) and \(W^{*}_i(\tau)\) are continuously differentiable functions.

Bagwell and Staiger (2005) assume that tariff changes generate standard terms-of-trade effects: the home country receives a (bilateral) terms-of-trade gain when it increases its tariff \(\tau_i\) on exports from foreign country \(i\), foreign country \(i\) experiences a terms-of-trade gain when it increases the tariff \(\tau_{ji}'\) that it applies to exports from the home country, and foreign country \(j\) for \(j \neq i\) enjoys a terms-of-trade gain when either \(\tau_i\) or \(\tau_{ji}'\) increases. The first two assumptions simply mean that each country is large. To motivate the third assumption, we note that an increase in \(\tau_i\) naturally directs home-country import demand toward foreign country \(j\)’s exports while an increase in \(\tau_{ji}'\) naturally reduces world demand for foreign country \(j\)’s import good.

As a general matter, tariff changes induce both local- and world-price effects; thus, the welfare effects of a given tariff change are not fully determined by the associated terms-of-trade effects. In view of the local- and world-price effects of changes in tariffs, Bagwell and Staiger (2005) thus do not impose general restrictions on the relationships between tariffs and reduced-form country welfare functions. Instead, they impose some additional structure on these relationships when tariffs are efficient, where efficiency is evaluated relative to country welfare functions. Specifically, for efficient tariffs, they assume that

\[
\begin{align*}
\frac{\partial W_i}{\partial \tau_i} &> 0 \quad \text{and} \quad \frac{\partial W^{*}_i}{\partial \tau_i} > 0, \\
\frac{\partial W_j}{\partial \tau_{ji}} &< 0 \quad \text{and} \quad \frac{\partial W^{*}_j}{\partial \tau_{ji}} < 0, \\
\frac{\partial W^*_i}{\partial \tau_i} &> 0 \quad \text{and} \quad \frac{\partial W^*_j}{\partial \tau_{ji}} > 0
\end{align*}
\]

for \(i, j = 1, 2\) and \(i \neq j\). In other words, when tariffs are efficient, Bagwell and Staiger (2005) assume that the welfare effects of a small change in any tariff are aligned with the corresponding terms-of-trade effects of that tariff change.

For the model considered by Bagwell and Staiger (2005), the assumption that \(\partial W_i/\partial \tau_i < 0\) ensures that foreign country \(i\) suffers a welfare loss from an externally generated terms-of-trade loss. As they observe, it then follows that the inequalities in the third line of (1) are in fact implied; that is, \(\partial W_j/\partial \tau_{ji} < 0\) implies \(\partial W_i/\partial \tau_i > 0\) and \(\partial W^*_j/\partial \tau_{ji} > 0\), since reductions in \(\tau_i\) and increases in \(\tau_{ji}'\) simply represent alternative external policy changes that induce a terms-of-trade loss for foreign country \(i\).

Under the assumptions given in (1), Bagwell and Staiger (2005) show that, at any efficient tariff vector, and for \(i, j = 1, 2\) and \(i \neq j\),

\[
-\frac{\partial W_i}{\partial \tau_i} \cdot \frac{\partial W_j}{\partial \tau_{ji}} > \frac{\partial W^*_i}{\partial \tau_i} / \frac{\partial W^*_j}{\partial \tau_{ji}} > 0 > -\frac{\partial W^*_i}{\partial \tau_i} / \frac{\partial W^*_j}{\partial \tau_{ji}}.
\]

This means that, at any efficient tariff vector, the home country and foreign country \(i\) could lower \(\tau_i\) and \(\tau_{ji}'\) in such a fashion as to enjoy mutual gains while imposing a welfare loss on foreign country \(j\). In effect, starting at any efficient tariff vector, the home country and foreign country \(i\) can move \(\tau_i\) and \(\tau_{ji}'\) 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.\(^{14}\)

\(^{13}\) At a more operational level, we may interpret the home-country tariffs as being discriminatory when \(\tau_i > \tau_{ji}'\) and the foreign exports belong to a common product classification.

\(^{14}\) This result is stated in Proposition 4 of Bagwell and Staiger (2005).
endogenous determination of the tariff vector \( \tau = (\tau^1, \tau^2, \tau^0) \) through bilateral negotiations.

Consider the bilateral negotiation between the home country and foreign country \( i \). Beginning from their initial tariffs \( \tau^0_1 \) and \( \tau^0_2 \) and taking \( \tau^1 \) and \( \tau^2 \) as given, the home country and foreign country \( i \) choose their Nash bargaining solutions to solve

\[
\max_{(\tau^1, \tau^2) \in \mathbb{R}^2} \Delta W^1(\tau^1, \tau^1, \tau^2, \tau^2, \tau^1_0, \tau^2_0) \cdot \Delta W^2(\tau^1, \tau^1, \tau^2, \tau^2, \tau^1_0, \tau^2_0) \\
\text{subject to}
\]

\[
W(\tau^1, \tau^1, \tau^2, \tau^2) \geq W(\tau^1_0, \tau^2_0, \tau^1_0, \tau^2_0),
\]

where \( S = \{x, y\} \subset \mathbb{R}^2 \) and \( 0 < x < y \).

\[
\Delta W^1(\tau^1, \tau^1, \tau^2, \tau^2, \tau^1_0, \tau^2_0) = [W(\tau^1, \tau^1, \tau^2, \tau^2) - W(\tau^1_0, \tau^2_0, \tau^1_0, \tau^2_0)]^\alpha
\]

and

\[
\Delta W^2(\tau^1, \tau^1, \tau^2, \tau^2, \tau^1_0, \tau^2_0) = [W^2(\tau^1, \tau^1, \tau^2, \tau^2) - W^2(\tau^1_0, \tau^2_0, \tau^1_0, \tau^2_0)]^\alpha.
\]

The bilateral negotiation between the home country and foreign country \( j \) is analogous. Beginning from their initial tariffs \( \tau^0_j \) and \( \tau^2_j \) and taking \( \tau^1_j \) as given, the home country and foreign country \( j \) choose their Nash bargaining solutions to solve

\[
\max_{(\tau^1_j, \tau^2_j) \in \mathbb{R}^2} \Delta W^1(\tau^1_j, \tau^1_j, \tau^2_j, \tau^2_j, \tau^1_0, \tau^2_0) \cdot \Delta W^2(\tau^1_j, \tau^1_j, \tau^2_j, \tau^2_j, \tau^1_0, \tau^2_0).
\]

subject to

\[
W(\tau^1_j, \tau^1_j, \tau^2_j, \tau^2_j) \geq W(\tau^1_0, \tau^2_0, \tau^1_0, \tau^2_0),
\]

where

\[
\Delta W^1(\tau^1_j, \tau^1_j, \tau^2_j, \tau^2_j, \tau^1_0, \tau^2_0) = [W(\tau^1_j, \tau^1_j, \tau^2_j, \tau^2_j) - W(\tau^1_0, \tau^2_0, \tau^1_0, \tau^2_0)]^\alpha
\]

and

\[
\Delta W^2(\tau^1_j, \tau^1_j, \tau^2_j, \tau^2_j, \tau^1_0, \tau^2_0) = [W^2(\tau^1_j, \tau^1_j, \tau^2_j, \tau^2_j) - W^2(\tau^1_0, \tau^2_0, \tau^1_0, \tau^2_0)]^\alpha.
\]

We may understand the inequality constraints in the respective programs as participation constraints. As captured by these constraints, for a given bilateral negotiation between the home country and foreign country \( i \), where \( i = 1, 2 \), if the negotiation results in disagreement, then the home country and foreign country \( i \) revert to the disagreement tariff pair \( (\tau^1_0, \tau^2_0) \) for their bilateral relationship. Importantly, the home country and foreign country \( i \) negotiate under the assumption that the “other” bilateral negotiation (i.e., the bilateral negotiation between the home country and foreign country \( j \), where \( j = 1, 2 \) and \( j \neq i \)) delivers the tariff pair \( (\tau^1_j, \tau^2_j) \), whether the bilateral negotiation between the home country and foreign country \( i \) results in agreement or disagreement.

Given \( S = \{x, y\} \subset \mathbb{R}^2 \) and \( 0 < x < y \), and for \(( \tau^1_0, \tau^2_0, \tau^1_0, \tau^2_0 ) \in S^4 \), for \(( \tau^1_0, \tau^2_0, \tau^1_0, \tau^2_0 ) \in S^4 \) is a Horn–Wolinsky solution if \(( \tau^1_0, \tau^2_0 ) \) solves (3) given \(( \tau^1, \tau^2 ) = ( \tau^1_0, \tau^2_0 ) \) and if \(( \tau^1_0, \tau^2_0 ) \).
solves (4) given \((\tau^1, \tau^{-1}) = (\tau_{hw}^1, \tau_{hw}^{-1})\). In other words, \(\tau_{hw} = (\tau_{hw}^1, \tau_{hw}^{-1}, \tau_{hw}^2, \tau_{hw}^{-2}) \in S^2\) 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 so as well.\(^{15}\)

We next define an interior Horn–Wolinsky solution as a Horn–Wolinsky solution for which \(\tau_{hw} = (\tau_{hw}^1, \tau_{hw}^{-1}, \tau_{hw}^2, \tau_{hw}^{-2}) \in (\bar{\tau}, \bar{\tau})^4\) and the following optimization conditions are satisfied:

\[
\frac{\partial W}{\partial \tau^i}/\frac{\partial \tau^i}{\partial \tau^j} = -\frac{\partial W^{-i}}{\partial \tau^i}/\frac{\partial \tau^i}{\partial \tau^j}, \text{ for } i, j = 1, 2.
\]

where all derivatives are evaluated at \(\tau_{hw}\). The optimization conditions in (5) are implied by the first-order conditions for the optimization of programs (3) and (4) under the following sufficient conditions: \(\tau_{hw} \in (\bar{\tau}, \bar{\tau})^4\), \(\partial W/\partial \tau^i, \partial W/\partial \tau^j\), and \(\partial W/\partial \tau^i, \partial W^{-i}/\partial \tau^j\) and \(\partial W^{-i}/\partial \tau^i\) are non-zero at \(\tau_{hw}\), and the participation constraints hold with slack at \(\tau_{hw}\).\(^{16}\) As we discuss in more detail in the next section, an interior Horn–Wolinsky solution thus ensures that the tariff pair agreed upon in any bilateral negotiation is “bilaterally efficient.”

4. 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 our model with simultaneous bilateral bargaining delivers an outcome, \(\tau_{hw} = (\tau_{hw}^1, \tau_{hw}^{-1}, \tau_{hw}^2, \tau_{hw}^{-2})\), where \(\tau_{hw}\) 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:

\[
\frac{\partial W}{\partial \tau^i}/\frac{\partial \tau^i}{\partial \tau^j} \geq 0 \text{ and } \frac{\partial W^{-i}}{\partial \tau^i}/\frac{\partial \tau^i}{\partial \tau^j} \geq 0, \text{ for } i, j = 1, 2.
\]

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

\[
\frac{\partial W}{\partial \tau^i} > 0 \text{ and } \frac{\partial W^{-i}}{\partial \tau^i} > 0, \text{ if } i, j = 1, 2, \text{ and } i \neq j
\]

\[
\frac{\partial W}{\partial \tau^i} < 0 \text{ and } \frac{\partial W^{-i}}{\partial \tau^i} < 0, \text{ if } \tau^i \text{ is a Nash-Wolinsky tariff vector}
\]

\[
\frac{\partial W^{-i}}{\partial \tau^i} > 0 \text{ and } \frac{\partial W^{-i}}{\partial \tau^i} < 0, \text{ if } \tau^i \text{ is an interior Nash-Wolinsky tariff vector}
\]

\(^{15}\) In the Appendix, we further develop the Nash-in-Nash representation of the Horn–Wolinsky solution and show that the solution can be interpreted as a generalized Nash equilibrium for a generalized two-person game in which the objective of player \(i\) is to choose \(\tau^i\) and \(\tau^{-i}\) so as to maximize the Nash bargaining solution objective \(\partial W^{-i}/\partial \tau^i\), while satisfying the associated participation constraints for the bargaining negotiation between the home country and foreign country \(i\). The game and solution concept are “generalized, since, due to the participation constraints, player \(i\)'s feasible strategy set is affected by the strategic choices of player \(j\). For \(i, j = 1, 2, \text{ and } i \neq j\). Given this representation, we can then utilize Debreu's (1952, 1953) existence theorem and directly provide sufficient conditions for the existence of a Horn–Wolinsky solution. See also Dasgupta and Maskin (2015) for further discussion of Debreu's contribution.

\(^{16}\) Formally, the participation constraints hold with slack at \(\tau_{hw}\) if \(W(\tau_{hw}) > \max \{W(\tau_{hw}^1, \tau_{hw}^{-1}, \tau_{hw}^2, \tau_{hw}^{-2}), W(\tau_{hw}^1, \tau_{hw}^{-1}, \tau_{hw}^3, \tau_{hw}^{-3})\}\). \(W^{-1}(\tau_{hw}) > W^{-1}(\tau_{hw}^1, \tau_{hw}^2, \tau_{hw}^{-1}, \tau_{hw}^{-2})\) and \(W^{-2}(\tau_{hw}) > W^{-2}(\tau_{hw}^1, \tau_{hw}^2, \tau_{hw}^{-1}, \tau_{hw}^{-2})\).

Under these assumptions, each country would like to increase its own tariff, each country does not want its export good to confront a higher tariff, and each foreign country \(i\) gains from an increase in either of the tariffs \(\tau^i\) and \(\tau^{-i}\) in the other bilateral trading relationship.

These assumptions can be interpreted in analogous fashion to the interpretation of (1) in Section 2. We recall that this interpretation is developed in terms of the first-scenario (homogeneous goods) model, as considered by Bagwell and Staiger (2005), but that similar properties would be expected in the second-scenario (differentiated goods) model as well, at least if the relevant products are sufficiently close substitutes. Further, and in line with our preceding discussion in Section 2, for the model considered by Bagwell and Staiger (2005), the assumption \(\partial W^{-i}/\partial \tau^i < 0\) ensures that foreign country \(i\) experiences a welfare reduction from an externally generated terms-of-trade loss and therefore implies that \(\partial W^{-i}/\partial \tau^i > 0\) and \(\partial W^{-i}/\partial \tau^i > 0\). In other words, the inequalities in the third line of (7) are then in fact implied by the second inequality in the second line of (7).

Starting at any such Horn–Wolinsky solution as captured by (6), and under the assumptions given in (7), 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 builds from footnote 11 in Bagwell and Staiger (2005). Bagwell and Staiger (2005) consider an efficient tariff vector and suppose that the tangency condition in (6) holds between the home country and some foreign country \(i\). They then consider a two-step perturbation as illustrated in Figure 2. In the first step, they increase \(\tau^i\) and \(\tau^{-i}\) in a fashion that maintains \(W^{-i}\). This corresponds to the movement from point A to point B in the figure. This first-step perturbation results in no change in \(W^{-i}\), a first-order increase in \(W^i\) and a second-order loss in \(W\) (due to the tangency between the isowelfare curves of the home country and foreign country \(i\)). The second step is then to increase \(\tau^i\) and decrease \(\tau^{-i}\) in a fashion that maintains \(W^{-i}\). We illustrate these steps in the figure with the movement from point C to point D. This second-step perturbation results in no change in \(W^{-i}\), a first-order loss in \(W^i\) and a first-order gain in \(W\). If the second-step perturbation is small relative to the first-step perturbation, then the perturbation in total results in no change in \(W^{-i}\) and first-order gains in \(W^i\) and \(W\), which contradicts the original hypothesis of an efficient tariff vector.\(^{17}\)

We consider here a similar perturbation but with three differences. First, we start with a situation in which the tangency condition (6) holds between the home country and both foreign countries. (By contrast, in the Bagwell–Staiger, 2005 perturbation just defined, the tangency condition (6) is assumed to hold between the home country and some foreign country \(i\)). Second, we want to find a perturbation...
that generates strict welfare gains for each of the three countries. (By contrast, in the Bagwell-Staiger, 2005 perturbation just defined, \(W^i\) is unchanged.) Third, we want to construct a perturbation under which all four tariffs are increased. (By contrast, in the Bagwell-Staiger, 2005 perturbation just described, \(\tau^{fi}\) is decreased.)

The key idea is to do two Bagwell-Staiger (2005) perturbations simultaneously, so that each foreign country plays the role 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 \(j\).

We now develop a formal representation of this idea. Specifically, starting at a tariff vector that satisfies (6), and under the assumption (7), we consider the following perturbation:

\[
dr^1 = dr^2 = \varepsilon + \sigma
\]

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-\(i\) and Home-\(j\) perturbations (with the designated foreign country playing the role of foreign country \(i\) in Bagwell and Staiger, 2005). The \(\varepsilon\) part of \(dW^i\) is then the gain in \(W^i\) from the Home-\(i\) step-1 increase in \(\tau^1\) and \(\tau^2\), where there is no first-order effect on \(W^i\) from the Home-\(i\) step-1 increase in \(\tau^1\) and \(\tau^2\). The \(\sigma\) part of \(dW^i\) is then the loss in \(W^i\) from the Home-\(i\) step-2 increase in \(\tau^1\) and decrease in \(\tau^2\) to keep \(W^i\) fixed, where by construction there is no effect on \(W^i\) from the Home-\(i\) step-2 increase in \(\tau^2\) and decrease in \(\tau^1\) that keeps \(W^i\) fixed.

Under our assumption that the initial tariff vector satisfies (7), the term in \(dW^1\) that is multiplied by \(\varepsilon\) is positive while the term that is multiplied by \(\sigma\) is negative; therefore, if \(\varepsilon\) is large relative to \(\sigma\) in the specific sense that

\[
e > \max_{(i,j=1,2,4)} \left\{ \frac{-\partial W^i}{\partial \tau^1} + \frac{-\partial W^j}{\partial \tau^1} - \frac{-\partial W^i}{\partial \tau^2} + \frac{-\partial W^j}{\partial \tau^2} \right\} \sigma
\]

then \(dW^1 > 0\). An exactly symmetric argument holds for foreign country \(j\).

Allowing for \(i = 1, 2\), we thus select \(\varepsilon > 0\) and \(\sigma > 0\) such that

\[
e > \max_{(i,j=1,2,4)} \left\{ \frac{-\partial W^i}{\partial \tau^1} + \frac{-\partial W^i}{\partial \tau^1} - \frac{-\partial W^j}{\partial \tau^2} + \frac{-\partial W^j}{\partial \tau^2} \right\} \sigma
\]

Under (11), 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 (8) that \(dr^1 = dr^2 > 0\). Referring to (9) and (10), we see for \(dr^i\) that the coefficient on \(\varepsilon\) is positive while that on \(\sigma\) is negative. Thus, we have that \(dr^i > 0\) for \(i, j = 1, \ldots, 4\).
2, \(i \neq j\) if and only if
\[
\varepsilon > \max_{i,j=1,2,\, i \neq j} \left( \frac{\partial W^{j}}{\partial \tau^{j}} \frac{\partial \tau^{j}}{\partial \tau^{i}} \right) \alpha. \tag{12}
\]

We now summarize our finding in the following proposition:

**Proposition 1.** Suppose the model with simultaneous bilateral bargaining delivers an interior Horn–Wolinsky solution, \(\tau_{h} = (\tau_{1h}, \ldots, \tau_{Nh})\), captured by (6). Suppose at this tariff vector that (7) holds. Then there exists a perturbation under which all tariffs are raised and all three countries enjoy strict welfare gains; in particular, for sufficiently small \(\varepsilon > 0\), \(\varepsilon > 0\) satisfying (11) and (12), the perturbation \(d\tau = (d\tau^{1}, d\tau^{2}, d\tau, d\tau^{2})\) defined in (8)–(10) is such that, for \(i = 1, 2\), \(d\tau^{1} > 0\), \(d\tau^{2} > 0\), \(d\tau^{2} > 0\), and \(dW^{i} > 0\).

We note that welfare gains accrue to all countries without separately assuming that (12) holds. The role of (12) is simply to ensure that all tariffs are increased as part of the perturbation.\(^{18}\)

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 all raise their tariffs. We thus have formalized an interpretation in which tariffs are inefficient in the sense of being too low, at any interior Horn–Wolinsky solution.

5. 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 small perturbation, then the perturbation cannot be characterized by “opportunistic” bilateral tariff changes in both bilateral relationships. Building from this finding, 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. An implication of our analysis is thus that there is no small perturbation from the Horn–Wolinsky solution that would make every country better off by lowering the tariffs of every country. In this way, 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.

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 (13) holds for \(i, j = 1, 2\) with \(i \neq j\). 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 (13) 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 \(i\):

\[
dW^{i} = \left( \frac{\partial W^{i}}{\partial \tau^{1}} + \frac{\partial W^{i}}{\partial \tau^{2}} \right) d\tau^{1} + \left( \frac{\partial W^{i}}{\partial \tau^{1}} + \frac{\partial W^{i}}{\partial \tau^{2}} \right) d\tau^{2} \geq 0,
\tag{13}
\]

where the inequality follows from the assumption that the welfare change is non-negative for all countries. Under (13), we see that the first term in this expression is negative; thus, it follows that

\[
\left( \frac{\partial W^{i}}{\partial \tau^{1}} + \frac{\partial W^{i}}{\partial \tau^{2}} \right) d\tau^{1} > 0.
\]

18 For the first-scenario model considered by Bagwell and Staiger (2005), and for \(i, j = 1, 2\) with \(i \neq j\), a perturbation entails an opportunistic bilateral tariff change in the Home–\(j\) bilateral relationship if and only if the bilateral tariff change described by \((d\tau^{1}, d\tau^{2})\) generates a terms-of-trade loss for foreign country \(i\). This follows since, starting at an interior Horn–Wolinsky solution, an externally generated terms-of-trade loss reduces \(W^{i}\) under (7).
Using $\partial W^i/\partial \tau^i < 0$ under (7), we may rewrite this inequality equivalently as
\[
\left[ \frac{\partial W^i}{\partial \tau^i} \frac{\partial W^i}{\partial \tau^j} + \frac{\partial \tau^i}{\partial \tau^j} \frac{\partial \tau^i}{\partial \tau^j} \right] \partial \tau^j < 0,
\]
where the inequality in (14) 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
\[
dW = \left[ \frac{\partial \tilde W}{\partial \tau^j} \frac{\partial \tilde W}{\partial \tau^i} + \frac{\partial \tau^i}{\partial \tau^j} \frac{\partial \tau^i}{\partial \tau^j} \right] \partial \tau^j \frac{\partial \tilde W}{\partial \tilde \tau^i} + \left[ \frac{\partial W^i}{\partial \tau^j} \frac{\partial W^i}{\partial \tau^i} + \frac{\partial \tau^i}{\partial \tau^j} \frac{\partial \tau^i}{\partial \tau^j} \right] \partial \tau^j \frac{\partial W^i}{\partial \tilde \tau^i} \partial \tilde W^j.
\]
where we use $\partial W/\partial \tau^j > 0$ for $i = 1, 2$ by (7). We now use the fact that an interior Horn–Wolinsky solution is bilaterally efficient and thus characterized by tangencies in each bilateral relationship. In particular, using (6), we now have that
\[
dW = \left[ \frac{\partial W^i}{\partial \tau^j} \frac{\partial W^i}{\partial \tau^j} \partial \tau^j \right] \partial \tilde W^j.
\]
where the inequality follows since each term in curly brackets is negative by (14) and $\partial W/\partial \tau^j > 0$ for $i = 1, 2$ by (7). 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 2.** Suppose the model with simultaneous bilateral bargaining delivers an interior Horn–Wolinsky solution, $\tau_{iaw} = (\tau^i_{iaw}, \tau^j_{iaw}, \tau^{ij}_{iaw}, \tau^{ij}_{iaw})$, captured by (6). Suppose at this tariff vector that (7) holds. Starting at this solution, a small perturbation $d\tau = (d\tau^i, d\tau^{ij}, d\tau^{ij}, d\tau^{ij})$ 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.

Proposition 2 identifies a key necessary feature for Pareto-improving perturbations. This result is of interest in its own right, but it also provides a stepping stone toward understanding the necessary features of tariff changes that deliver Pareto gains. We thus conclude this section by exploring the implications of this proposition for the nature of the underlying tariff changes that a small, Pareto-improving perturbation must deliver.

To this end, let us fix an interior Horn–Wolinsky solution, $\tau_{iaw} = (\tau^i_{iaw}, \tau^j_{iaw}, \tau^{ij}_{iaw}, \tau^{ij}_{iaw})$, captured by (6). For $i, j = 1, 2$ and $i \neq j$, given that (7) holds at $\tau_{iaw}$, we know that foreign country $j$ would gain from a small increase in either of the tariffs $\tau^i$ and $\tau^{ij}$ in the Home–$j$ bilateral relationship; thus, in a graph with $\tau^i$ and $\tau^{ij}$ on the axes, the isowelfare curve for foreign country $j$ takes a negative slope as it passes through the Horn–Wolinsky tariff pair $(\tau^i_{iaw}, \tau^{ij}_{iaw})$. In the neighborhood of $(\tau^i_{iaw}, \tau^{ij}_{iaw})$, it follows, too, that tariff pairs above (below) foreign country $j$’s isowelfare curve generate higher (lower) welfare for foreign country $i$. Hence, if starting at $\tau_{iaw}$, we were to consider a small perturbation $d\tau = (d\tau^i, d\tau^{ij}, d\tau^{ij}, d\tau^{ij})$ such that the bilateral tariff change in the Home–$j$ bilateral relationship is not opportunistic, then the bilateral tariff change described by $(d\tau^i, d\tau^{ij})$ must generate a tariff pair that lies on or above the described isowelfare curve for foreign country $i$. Accordingly, for such a perturbation, it is necessary that $d\tau^i \geq 0$ or $d\tau^{ij} \geq 0$.

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 strictly gains under the perturbation. The latter property rules out the trivial possibility where all tariffs are unaltered.

Consider a small perturbation satisfying these properties. A first point is that, for all $j = 1, 2$, the perturbation must entail a change in $\tau^j$ or $\tau^{ij}$. To see why, suppose that the perturbation changes neither $\tau^j$ nor $\tau^{ij}$. Now consider the Home–$i$ bilateral relationship, where $i = 1, 2$ and $i \neq j$. Given that, at an interior Horn–Wolinsky solution, $\tau^i$ and $\tau^{ij}$ are set in a bilaterally efficient manner for the Home–$i$ bilateral relationship, any change in $\tau^j$ and $\tau^{ij}$ must result in a welfare loss for the home country or foreign country $i$. Furthermore, if $\tau^j$ and $\tau^{ij}$ were also unaltered, then the perturbation would fail to generate a welfare change for any country. We conclude that, for all $j = 1, 2$, the assumed properties of the perturbation necessitate a change in $\tau^j$ or $\tau^{ij}$. A second point, which follows from **Proposition 2**, is that the bilateral tariff change in at least one bilateral relationship is not opportunistic. Suppose that the Home–$j$ bilateral relationship is not opportunistic. As we argue just above, the bilateral tariff change described by $(d\tau^i, d\tau^{ij})$ then must satisfy $d\tau^{ij} \geq 0$ or $d\tau^{ij} \geq 0$. Given that our first point rules out $d\tau^j = 0$ or $d\tau^{ij} = 0$, it follows that $d\tau^j > 0$ or $d\tau^{ij} > 0$.

We may thus conclude the section with the following proposition:

**Proposition 3.** Suppose the model with simultaneous bilateral bargaining delivers an interior Horn–Wolinsky solution, $\tau_{iaw} = (\tau^i_{iaw}, \tau^j_{iaw}, \tau^{ij}_{iaw}, \tau^{ij}_{iaw})$, captured by (6). Suppose at this tariff vector that (7) holds. Starting at this solution, a small perturbation $d\tau = (d\tau^i, d\tau^{ij}, d\tau^{ij}, d\tau^{ij})$ generates a strict welfare gain for at least one country while not lowering the welfare of any other country only if at least one tariff rises; that is, starting at this solution and for a small perturbation $d\tau$, if for $i = 1, 2$, $d\tau^j > 0$ and $d\tau^{ij} > 0$ with at least one inequality strict, then there exists $j \in [1, 2]$ such that $d\tau^j > 0$ or $d\tau^{ij} > 0$.

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, we know that a weak Pareto improvement requires that, in at least one bilateral relationship, the associated bilateral tariff changes generate a weak welfare gain for the non-member foreign country. As we argue above, if we assume further that the perturbation generates a strict welfare gain for at least one country, then we can conclude that at least one tariff in this bilateral relationship rises. The other tariff in this bilateral relationship may rise as well or

---

20 We use “or” in the inclusive sense; thus, “$x$ or $y$” allows that $x$ and $y$ may be both true.
it could fall. But if it falls, it cannot fall to such an extent as to reverse the weak welfare gain that the non-member foreign country must enjoy.

We now offer a simple description of the implications of our findings for this and the previous section: Starting at an interior Horn–Wolinsky solution, it is not possible to make every country better off with a small perturbation that induces a weak reduction in the tariffs of every country (Proposition 3), but it is possible to make every country better off with a perturbation that generates a small increase in the tariffs of every country (Proposition 1). Together, these implications provide reinforcing support for our central message that tariffs are inefficiently low in an interior Horn–Wolinsky solution.21

6. Conclusion

We consider a three-country model of simultaneous bilateral tariff negotiations where each country is affected by the outcomes achieved in each bilateral negotiation. Allowing for discriminatory tariffs, we characterize the negotiated tariffs that are predicted by the Horn–Wolinsky (1988) solution. We show that starting from an interior Horn–Wolinsky solution we can construct a perturbation under which all tariffs are increased in a way that generates welfare gains for all countries. We also characterize the necessary features of small, Pareto-improving perturbations, showing 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 short, starting at an interior Horn–Wolinsky solution, it is not possible to make every country better off with a small perturbation that induces a weak reduction in the tariffs of every country, but it is possible to make every country better off with a perturbation that generates a small increase in the tariffs of every country. Based on these findings, we conclude that simultaneous bilateral tariff negotiations are associated with excessive liberalization when judged relative to the preferences of all countries.

Our work also contributes at a methodological level. 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, such as the quantitative trade model studied in Bagwell et al. (2018a). Our results also may be useful for applications in other fields, particularly as our assumptions are placed on general reduced-form welfare functions that map negotiation outcomes to payoffs. In addition, 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.

Throughout the paper, we allow countries to negotiate discriminatory tariffs. An interesting direction for future research is to study the application of the Horn–Wolinsky solution concept to simultaneous bilateral tariff negotiations when tariffs must be non-discriminatory (i.e., satisfy the MFN rule). The logic of the Horn–Wolinsky solution concept extends in straightforward fashion to settings with simultaneous bilateral bargaining under the MFN rule when each country imports multiple goods if each country negotiates its tariff for any given import good only with a single principal supplier of that good. Some conceptual challenges arise in applying the Horn–Wolinsky solution, however, if a country negotiates its MFN tariff on a given good simultaneously with multiple partners. In particular, for the three-country model that we consider here, if the home country simultaneously negotiates with both foreign countries over its MFN tariff, then the following questions must be addressed: Which negotiated MFN tariff for the home country is ultimately applied, and do all participants understand the process through which this determination is made at the time of their respective negotiations? We leave a thorough analysis of these and other questions to future research.

Acknowledgements

For helpful comments, we thank Arnaud Costinot, Henrik Horn, two anonymous referees and seminar participants at LSE, Stanford University and the 2017 Research Workshop on the Economics of International Trade Agreements at Villars. We thank the National Science Foundation (Grant SES-1326940) for financial support. We thank Ohyun Kwon for excellent research assistance. The preceding version of this paper (Bagwell et al., 2018) was circulated with the title “Nash-in-Nash” Tariffs with and without MFN.

Appendix

To establish conditions for the existence of a Horn–Wolinsky solution as defined in the text, we define a generalized game with infinite strategy spaces and two players. The objective of player $i$, where $i = 1, 2$, is to select $\tau_i$ and $\tau_i^+$ so as to maximize the Nash bargaining solution objective for the bargaining relationship between the home country and foreign country $i$. Each player $i$, however, must also select $\tau_i^-$ and $\tau_i^+$ from the space of feasible tariffs that satisfy participation constraints, as captured by the weak inequalities stated in the text. We note that the participation constraints for player $i$ are affected by the strategy choices of player $j$, where $j = 1, 2$ and $i \neq j$.

Formally, player $i = 1, 2$ has a strategy $s_i = (\tau_i, \tau_i^\pm)$, where $s_i \in S = \{\tau, \tau^\pm\}$ with $(\tau, \tau) \in \mathbb{R}^2$ and $0 < \tau, \tau^-$. Player $i$ has the payoff function $g_i(s_i, s_j)$, where

$$g^1(s_i, s_j) = \Delta W^1(s_i, s_j) = \Delta W^1(s_i, s_j)$
$$

and where $s_i = (\tau_i, \tau_i^0) \in S, i = 1, 2$, are exogenously given. Finally, in recognition of the participation constraints, we further restrict $s_i$ to a subset $\gamma'(s_i) \in S$, where we define

$$\gamma^1(s_i) = \{s_i \in S|W(s_i, s_j) \geq W(s_i, s_j) \text{ and } W(s_i, s_j) \geq W(s_i, s_j)\}
$$

$$\gamma^2(s_i) = \{s_i \in S|W(s_i, s_j) \geq W(s_i, s_j) \text{ and } W(s_i, s_j) \geq W(s_i, s_j)\}.
$$

We now say that a pair $(s_1, s_2)$ is a generalized Nash equilibrium if, for all $i, j = 1, 2$ with $i \neq j$, $s_i \in \gamma'(s_i)$, $g_i(s_i, s_j) \geq g_j(s_i, s_j)$ for all $s_j \in \gamma'(s_i)$, and $g_i(s_i, s_j) \geq g_j(s_i, s_j)$ for all $s_j \in \gamma'(s_i)$. The Horn–Wolinsky solution may now be understood as a generalized Nash equilibrium for the two-person generalized game defined here.

We recall that $S = \{s, s^0\}$ and that $W(\tau, \tau)$ and $W^{\pm}(\tau, \tau)$ are continuously differentiable for $i = 1, 2$. It follows that $S$ is a nonempty, compact and convex subset of Euclidean space and that, for $i = 1, 2, g_i(s_i, s_j)$ is continuous in $(s_i, s^0)$. According to Debreu’s (1952, 1983) theorem, a pure strategy generalized Nash equilibrium exists for the generalized two-person game defined here if for $i, j = 1, 2$ and $i \neq j$, $g_i(s_i, s_j)$ is quasiconcave in $s_i$ and $b_i(s_i) is upper and lower hemicontinuous, convex valued and nonempty valued. Equivalently, conditions (a) and (b) ensure the existence of a Horn–Wolinsky solution for the model defined in the text.

21 In our Online Appendix, we also explore a particular representation of our model so as to concretely illustrate and further develop the themes described above. The representation that we consider is a two-good endowment economy in which consumers have Cobb-Douglas preferences that weigh both goods equally. Under the assumption that each government maximizes the indirect utility of the representative agent in its country, we provide numerical characterizations of Nash tariffs, efficient tariffs, and the interior Horn–Wolinsky solution. Among other findings, we verify that the interior Horn–Wolinsky solution exists for this representation.

22 Some initial perspectives on these issues are found in our working paper (Bagwell et al., 2018).
Note that condition (a) imposes quasiconcavity in \((\tau^i, \tau^{i'})\) for the Nash Bargaining solution objective, \(\Delta W(\tau^1, \tau^{-1}, \tau^2, \tau^{-2}; \tau^l, \tau^r)\). We can show that condition (b) holds if, for each \(i, j = 1, 2\) with \(i \neq j\) and for any \(s \in \mathbb{S}\), \(W(\tau)\) and \(W^{i}(\tau)\) are strictly quasiconcave in \(s_i\). It is direct to verify that \(\gamma(s)\) is nonempty (since \(s_i\) is a member), convex valued (since \(W(\tau)\) and \(W^{i}(\tau)\) are quasiconcave in \(s_i\)) and upper hemicontinuous (since \(W(\tau)\) and \(W^{i}(\tau)\) are continuous). Using the strict quasiconcavity of \(W(\tau)\) and \(W^{i}(\tau)\) in \(s_i\), we can also show that \(\gamma(s)\) is lower hemicontinuous in our setting. Finally, we also note that the conditions stated here ensure existence but do not ensure interiority.

**Supplementary Material: Appendix A**

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jinteco.2019.103263.

**References**


---

23 In the supplementary materials provided in our Online Appendix, we consider a particular representation of the model as an endowment economy with Cobb-Douglas preferences. Under the assumption that each government maximizes the indirect utility of a representative agent in its country, we use a numerical example to illustrate that condition (a) plausibly holds for examples of interest.