3-2020

Sustainable cooperation in international trade: A quantitative analysis

Yuan MEI
Singapore Management University, yuanmei@smu.edu.sg

Citation
MEI, Yuan. Sustainable cooperation in international trade: A quantitative analysis. (2020). *Journal of International Economics*. 123, 1-17.
Available at: https://ink.library.smu.edu.sg/soe_research/2362

Sustainable Cooperation in International Trade: A Quantitative Analysis

Yuan Mei[*]

Singapore Management University

December 22, 2018

Abstract

How does the presence of multilateral institutions affect the sustainability of trade-policy cooperation? Do free-trade agreements make multilateral cooperation less sustainable? Will countries be more likely to deviate from negotiated tariffs when more trade liberalization realizes in the future? These questions have been studied in theory literature using models that feature repeated games, but have yet to be quantitatively analyzed. In this paper, I propose a methodology to quantitatively characterize the equilibrium strategies on tariffs of various nations in a widely used repeated-game framework. I then apply this methodology to address these questions from a quantitative perspective. The numerical results computed from a reasonably comprehensive general equilibrium trade model corroborate previous analysis derived theoretically from simpler trade models. However, only free-trade agreements appear to influence the sustainability of trade-policy cooperation with quantitative significance.

JEL Classification: C7, F1, F4

Keywords: propensity to deviate, quantitative trade policy, repeated game

---

[*] School of Economics, Singapore Management University, 90 Stamford Road, Singapore, 178903 (e-mail:*yuanmei@smu.edu.sg*). I am grateful to Ralph Ossa, Jonathan Dingel, Felix Tintelnot, Ayumu Kikkawa, Yunzhi Hu, and seminar discussants at University of Chicago and Midwest Trade Conference (Purdue) for very helpful comments and suggestions. All errors are my own.

I Introduction

The global trading system is one of the most prominent examples of successful international policy cooperation, as countries maintain liberalized trade despite the opportunity each has to gain by raising their import tariffs. To explain why low tariffs can be sustained, interactions in international trade policy are typically modeled as a repeated game. Because a country can improve welfare by imposing tariffs that are higher than negotiated levels, deviation can only be deterred when future loss as a result of deviation is larger than the gain. This framework has also been used to study how the existence of multilateral institutions such as the World Trade Organization (WTO), preferential trade agreements, or changes in trade volume can affect the sustainability of low tariffs.

Although existing theoretical analysis on trade policy has widely used repeated-game models, related quantitative works using the same framework is rare. If the observed low tariffs are considered an equilibrium outcome of the repeated game, each country's optimized gain from imposing higher tariffs is outweighed by subsequent welfare loss due to other countries' retaliation in the future. Without characterizing the equilibrium quantitatively, however, determining whether such low tariffs can still be sustained under external changes can be difficult. For example, when two countries sign a free-trade agreement (FTA), gains and future loss as a result of deviation will be different not only for the participating countries but for other countries as well. To investigate whether trade-policy cooperation is still sustainable requires quantifying how these welfare consequences change in the new equilibrium of the repeated game.

In this paper, I propose a methodology to quantitatively characterize countries' equilibrium strategies on tariffs. This methodology is compatible with numerous theoretical works that rely on repeated games to explain observed low tariffs, and can be used to study how changes in the world trading system can influence trade-policy cooperation. I first introduce the features these repeated-game models share. In particular, by describing each stage game with a static trade model, I can compute each country's one-period gain from deviation and the future loss as a result of the deviation. Given these computed welfare changes, I can then compute the minimum patience (i.e., discount factor) needed such that an infinite reversion to the punishment phase can sustain the cooperative tariffs. When the deviating country does not incur any loss in punishment phase (the case of winning a trade war first discussed in Johnson (1954), for example), this number will be larger than one, implying that low tariffs are not sustainable. When the minimum patience is less than one, we can still infer which country is more likely to deviate by comparing this measure across countries. For this reason, throughout the paper I will refer to this computed minimum discount factor as a measure of the propensity to deviate.

I then use a multi-country, multi-industry quantitative trade model to illustrate how this methodology can be applied. In this reasonably comprehensive general equilibrium model, governments can use tariffs to improve welfare, measured by either real expenditure or sectoral profits weighted by political incentives as in Grossman and Helpman (1994). Focusing on the major players in international trade policy negotiation, I first numerically compute each country's static welfare change in the deviation phase: each country imposes tariffs that maximize its objective function given other countries' tariffs remaining unchanged. Not surprisingly, all countries can improve their own welfare at the expense of other countries when deviating from low tariffs. In the benchmark case where welfare is measured by real expenditure, the potential one-period welfare gain ranges from 1.1% (Mercosur) to 4.5% (Canada).

I then compute static welfare changes in four counterfactual scenarios relevant to the equilibrium strategies in the punishment phase. Among them, the case of imposing Nash tariffs has been most commonly used in the theoretical analysis of trade policy cooperation, although the case of autarky has also been considered.[1] I find that when these two forms of punishments are used, the future loss of any deviating country outweighs the gain from deviation and the calculated minimum discount factor to sustain cooperative tariffs is very small. In other words, the negotiated low tariffs are sustainable when reverting to trade war or autarky infinitely are used to deter deviation. In addition, I also consider two other forms of punishment that are less stringent. In the case of the weakest punishment in which other countries impose WTO-bound tariffs against the deviating country, the stage equilibrium with low tariffs is not sustainable: the calculated minimum discount factor is greater than one for all countries.

Next, I present how this methodology can be applied to study theoretical findings in trade-policy literature that have yet to be quantitatively analyzed. I first examine whether multilateral institutions such as the WTO reduce the incentives of countries to deviate. Here I focus on the WTO's role of verifying deviation and facilitating multilateral enforcement efforts.[2] In other words, in a world without multilateral institutions, countries can impose deviating tariffs against one other and only be punished by the victim. Maggi (1999) uses a repeated-game framework that incorporates a three-country, partial-equilibrium trade model to show countries can, in theory, sustain a higher equilibrium payoff with multilateral enforcement than with bilateral enforcement. This is mostly because the WTO can introduce punitive tariffs from third parties, which increases potential loss from deviation. I compare the impact of the two enforcement mechanisms on the sustainability of tariffs using the more comprehensive general equilibrium trade model, but still under the same repeated-game framework. The quantitative results support Maggi's claim qualitatively: contrary to the case under the multilateral enforcement mechanism, the factual tariffs become unsustainable

---

[1] Examples of theoretical works involving infinite Nash reversion include Bagwell and Staiger (1990), Bagwell and Staiger (1999a), and Limao and Saggi (2008). Discussions about using autarky to deter deviations can be found in Dixit (1987), Bagwell and Staiger (1990) and Park (2000).

[2] Bagwell, Bown and Staiger (2016) offers a comprehensive literature review of the role of GATT/WTO.

under the bilateral enforcement mechanism as countries now gain by imposing deviating tariffs against their weakest trading partners. However, the welfare consequences of removing the WTO is in general not quantitatively significant.

I then turn to analyze whether preferential trade agreements promote or jeopardize the sustainability of multilateral trade cooperation. This question has attracted greater attention since the famous "building block" versus "stumbling block" argument raised by Bhagwati (1991). Although related theoretical research has been fruitful, quantitative work is rare and mostly focuses on how reductions in external most-favored-nation tariffs vary with exposure to preferential trade arrangements.[3] I first quantitatively identify the tariff complementarity effect and the punishment effect of enforceable bilateral FTAs. The tariff complementarity effect describes the situation in which an FTA will cause its member countries to impose lower external tariffs in the case of deviation. The punishment effect, on the other hand, is like the mirror image of the tariff complementarity effect: the FTA member countries will impose lower punitive tariffs against deviating non-member countries. According to Bagwell and Staiger (1999b), these two effects are the main channels through which FTAs can affect the sustainability of multilateral trade cooperation in the standard repeated game framework. The numerical results show these two effects on average reduce median optimal tariffs in the deviation phase and punishment phase by 12.7% and 11.6%, respectively. I then compare how the propensity of countries to deviate changes when FTAs are negotiated. I find that a new FTA always makes non-member countries more likely to deviate from negotiated tariffs, whereas the propensity to deviate could either increase or decrease for member countries.

Finally, I analyze the impact of global trade liberalization on sustainability of trade-policy negotiation. I model trade liberalization as an exogenous, symmetric reduction in tariffs for all 10 countries included to study how the propensity of countries to deviate varies along this process. Treating the reduced tariffs as negotiated tariffs to which countries agree to adhere, I can then apply the methodology developed previously to trace how countries' propensity to deviate varies along with worldwide trade liberalization. My results indicate that, with the exception of Japan, the propensities to deviate for all other countries are relatively stable.

Previous quantitative analysis of trade-policy negotiation that incorporates repeated games is rare, despite the framework's popularity in theoretical studies. My work is closely related to Ossa (2014), which provides a first numerical analysis of non-cooperative and cooperative trade policy. Relying on the same procedure to compute welfare changes in the deviation phase and punishment phase, I extend his work by adding inter-temporal interactions into the static trade model. Incorporating the repeated-game framework allows me to quantitatively investigate various theoretical findings from existing literature. For example, I believe this paper is the first attempt to quantify the tariff complementarity effect and the punishment effect of FTAs discussed in Bagwell and

---

[3] see Estevadeordal, Freund and Ornelas (2008) and Limao (2006), for example.

Staiger (1999b).[4] Likewise, no precedent exists for quantitatively identifying the extent to which multilateral institutions can improve the sustainability of trade negotiation as theorized in Maggi (1999).

Whereas theory literature mostly considers Nash tariffs as punishment once deviation is observed, I also consider autarky and two other means of punishment in the quantitative exercise. The first scenario is sanction: countries in each period of the punishment phase impose optimal tariffs against the deviating country while keeping cooperative tariffs among each other. In addition, I also consider the possibility that the punishment phase is still disciplined by WTO rules and bound tariffs are imposed against the deviating country. This punishment strategy is much weaker than the other three and all countries will choose to deviate from cooperative tariffs in the benchmark case. In other words, when a country wants to abandon WTO rules and deviate from low tariffs, punishments that are constrained by WTO are no longer effective deterrents. By introducing two more plausible means of punishment, this paper also relates to existing research on deterrents to deviating from cooperative tariffs and the role of WTO's dispute settlement.[5]

Although the static trade model I use in this paper builds on the one first discussed in Krugman (1980), the methodology introduced is flexible enough to incorporate alternative trade models. For example, I do not foresee any obstacle using other quantitative trade models such as those in Eaton and Kortum (2002) and Melitz (2003) to quantify equilibrium strategies. Ossa (2015) shows that gains from trade among these gravity models will differ once the industry dimension of trade flows is taken into account. Hence, choosing other static models that feature multiple sectors will lead to different characterizations. In addition, the procedure discussed in this paper should not have any problem incorporating other plausible means of punishment.

The data used in this paper build upon those analyzed in Ossa (2016). Throughout this paper, I focus on the world's nine largest economies[6] and a residual Rest of the World. Each economy is further disaggregated into 33 industries. Trade, production, and applied tariff data come from the Global Trade Analysis Project database (GTAP 8) for the year 2007. Following the "exact hat algebra" popularized by Dekle, Eaton and Kortum (2007), I express equilibrium conditions in changes to reduce the number of parameters that need to be calibrated. In addition, calculating tariffs imposed by welfare-maximizing governments in various counterfactual scenarios involves high dimensional numerical optimization. Hence, to improve efficiency, I use the method of mathematical programming with equilibrium constraints (MPEC), which is first introduced by Su and Judd (2012).

---

[4] Saggi, Stoyanov and Yildiz (2018) empirically confirms the existence of the complementarity effect. Their focus, however, is on MFN tariffs instead of the sustainability of trade policy cooperation.

[5] See Beshkar (2010) and Bown (2004), for example.

[6] These countries are Canada, China, India, Japan, Russia, South Korea, the United States, the EU-25 countries, and the Mercosur countries.

The remainder of this paper is organized as follows. The next section develops a methodology to quantitatively characterize countries' equilibrium strategies in a repeated-game framework, and provides some intuitions on why the minimum patience needed to sustain cooperative tariffs can be considered a measure of the incentive to deviate. Section III applies the methodology with a standard multi-sector trade model and presents some numerical results. In the subsequent three sections, I then study some theoretical findings in the existing literature from a quantitative perspective. Specifically, Section IV is devoted to analyzing the impact of multilateral institutions on the sustainability of trade-policy cooperation. Section V focuses on the impact of FTAs, and Section VI discusses how countries' propensity to deviate varies with exogenous worldwide trade liberalization. The last section offers some conclusions.

II Methodology

In this section, I develop a methodology to quantify equilibrium strategies of trade policy in a repeated-game framework. Subsection II.1 discusses the setup of the repeated game and static welfare changes that can be computed from quantitative trade models. In Subsection II.2, I characterize the equilibrium strategy that will be used to quantify the sustainability of trade-policy cooperation in following sections.

II.1 Repeated Game

Consider any static trade model in which the government's only control is tariffs. Hence, the only parameters that are allowed to change exogenously in the trade model are tariffs. We can let $W_j(t_j, t_{-j})$ denote the one-period welfare of the government in country j, which is determined by its own tariffs and all other countries' tariffs. Note $W_j(t_j, t_{-j})$ can either be the real expenditure that represents the households' welfare in country j, or some measures that capture political interests as in Grossman and Helpman (1994) and Ossa (2014). In each period, every country with discount factor $\beta \in (0, 1)$ sets its own tariffs, and current-period welfare is determined. In the next period, all past choices are observed and new tariffs are set. The stage game is repeated infinitely. This repeated game is stationary in the sense that none of the parameters change over time.

I assume the repeated game starts with negotiated tariffs all countries have agreed to follow. Although they resemble the notion of cooperative tariffs in trade-policy literature, negotiated tariffs do not take any stand on whether they are on the efficient frontier. In fact, I focus only on the sustainability of some given set of tariffs throughout this paper. Whether the negotiated tariffs maximize the joint welfare of all countries is beyond the scope of this paper. I follow the terminology used in game theory literature and define the equilibrium in the stage game in which no country deviates from the negotiated tariffs as the static cooperative equilibrium.

This repeated game has multiple equilibria. I will focus on those equilibria that are relevant to trade-policy negotiation. When a country chooses to deviate from negotiated tariffs, it will impose tariffs that maximize its static welfare, given that other countries' tariffs do not change. I define the deviation phase of this repeated game as the period in which one country deviates but the static cooperative equilibrium has been kept in all previous periods. Let superscript D and C denote optimal deviating tariffs in the deviation phase and negotiated tariffs in the cooperative equilibrium respectively. I can then define country j's (one-period) percentage welfare gain from deviation:

$$\Omega_j(t_j^D) \equiv \frac{W_j(t_j^D, t_{-j}^C)}{W_j(t_j^C, t_{-j}^C)} - 1. \tag{1}$$

Throughout this paper, I will focus only on cases in which Ω_j is positive. In other words, I assume every country can improve its welfare by deviating from negotiated tariffs. To sustain the static cooperative equilibrium, equilibrium strategies need to include a punishment phase to deter deviation from negotiated tariffs. The punishment phase is defined as the periods following the deviation phase in which other countries impose tariffs in such a way that the deviating country is expected to experience a welfare loss. The deviating country will still impose tariffs to maximize its own welfare given other countries' tariffs in the punishment phase, but such welfare-maximizing tariffs will be different from those in the deviation phase. On the other hand, other countries do not necessarily maximize their welfare in the punishment phase. Let superscript P denote tariffs in the punishment phase. I can also define country j's static one-period percentage welfare loss from deviation in the punishment phase:

$$\Phi_j(t_j^P, t_{-j}^P) \equiv \frac{W_j(t_j^P, t_{-j}^P)}{W_j(t_j^C, t_{-j}^C)} - 1. \tag{2}$$

Throughout this paper, I assume that $\Omega_j(t_j^D) > \Phi_j(t_j^P, t_{-j}^P)$. In other words, it is impossible for the deviating country's welfare to improve more in the punishment phase than in the deviation phase. When other countries are able to punish the deviating country at least to some extent, we should expect Φ_j to be negative.

A country's choice on tariffs in the deviation phase does not involve much ambiguity, because the deviating country will always choose tariffs that maximize static welfare. On the other hand, what tariffs should be imposed in the punishment phase to deter deviation is not as obvious. Existing theoretical research mostly considers two types of punishment tariffs. The more popular punishment involves a Nash equilibrium in which players simultaneously maximize their own welfare given others' tariffs. Moreover, as Dixit (1987) has pointed out, autarky is also a possible equilibrium in the stage game. Even though countries' intuitive understanding will serve to avoid autarky in the static game, threatening to revert to autarky can more effectively deter cheating.[7]

Whereas Nash tariffs and autarky are commonly used in the theoretical analysis of trade negotiations, they are generally considered extremely harsh punishments in reality. For example, imposing Nash tariffs in the punishment phase would lead to a "global trade war" scenario in trade models with more than two countries and would imply deviations from negotiated tariffs even among the non-deviating countries. In the numerical analysis of this paper, I introduce two more types of punishment that are less harsh than Nash tariffs and autarky. The first one is a bound tariff: the maximum rate of tariff allowed by the WTO to be imposed among WTO member countries. I also consider the possibility that when one country deviates from the cooperative equilibrium in the previous period in a repeated game, all other countries impose Nash tariffs to the deviating country only, but sustain negotiated tariffs between each other. Although these four scenarios in the punishment phase all act as deterrents against deviation from negotiated tariffs, they do lead to different quantitative results when the sustainability of the cooperative equilibrium is analyzed in later sections.

II.2 Equilibrium Strategy

Note that future welfare loss as a result of deviation depends not only on means of punishment (and hence static welfare loss), but also on equilibrium strategies of the repeated game. The one equilibrium strategy the trade-policy literature has studied intensively is that of keeping the cooperative stage equilibrium when no country deviates, and having a punishment phase of infinite periods if any deviation is observed. This strategy is formally defined as follows:

1. Each country imposes cooperative tariffs t^C in every stage game unless some country deviated from cooperation in the past.

2. If one country has not played t^C in the past and some other countries incur a welfare loss in the deviation phase, every country will enter a permanent punishment phase and impose t^P for all future periods.

To analyze the sustainability of the cooperative stage equilibrium, define $\underline{\beta_j}$ as the minimum patience (discount factor) needed to satisfy the self-enforcing constraint of country j:

$$\Omega_j(t_j^D) + \sum_{n=1}^{\infty} \underline{\beta_j}^n \Phi_j(t_j^P, t_{-j}^P) = 0.$$

---

[7] This argument is also raised in Park (2000) which shows that the autarky punishment instead of the interior Nash punishment may provide the small country with greater bargaining power.

We can then express $\underline{\beta_j}$ in terms of welfare changes in the deviation and punishment phase:

$$\underline{\beta_j} = \frac{\Omega_j(t_j^D)}{\Omega_j(t_j^D) - \Phi_j(t_j^P, t_{-j}^P)} \tag{3}$$

The above strategy constitutes a Nash equilibrium in which no country will deviate if $\beta > \underline{\beta_j}, \forall j$. Therefore, a lower $\underline{\beta_j}$ implies that country j is less likely to deviate from the cooperative stage equilibrium.[8] Note that when Φ_j is positive, the resulting $\underline{\beta_j}$ will be larger than one. Since $\beta \in (0,1)$, the cooperative stage equilibrium will be unsustainable. One possible explanation of the positive Φ_j is the situation described in Johnson (1954), in which a large country can actually win a bilateral trade war. In this case, because the deviating country does not incur a welfare loss in the punishment phase, the cooperative stage equilibrium cannot be sustained.

Using $\underline{\beta_j}$ to measure the sustainability of cooperative tariffs exploits one distinctive feature of the repeated game framework: the deviating country will face other countries' retaliation one period after the deviation. Whereas this framework is mostly commonly used in the theoretical analysis of trade policy cooperation, this feature of lagged punishment is actually supported by empirical evidence. The Dispute Settlement Body (DSB) of the WTO only allows retaliatory responses to start after the DSB has concluded the panel process and granted authorization. Horn, Johannesson and Mavroidis (2011) studies the 426 WTO disputes from 1995 to 2010. They find that the average panel process of the disputes is 445 days, even though the Dispute Settlement Understanding (DSU) of the WTO suggests that the completion of the panel process should not be more than nine months. Among all the disputes studied, the panel process had been completed within the statutory limits in only ten instances.

In Section B of the Appendix, I also discuss an alternative strategy that involves finite periods of punishment. If finite periods of punishment is considered, then the minimum length of the punishment phase needed to deter a country from deviating can also be used to measure the sustainability of the cooperative stage equilibrium. In the quantitative analysis in the following sections of this paper, I consider the equilibrium involving infinite punishment as the equilibrium strategy and use $\underline{\beta}$ to measure the sustainability of trade policy cooperation. This is mainly because infinite punishment is the most commonly used equilibrium strategy in existing theoretical analysis of trade policy cooperation.[9] In addition, computing $\underline{\beta}$ is straightforward and does not involve

---

[8] If country j experiences a temporary preference shock that reduces its discout factor to a value less than $\underline{\beta_j}$, sustaining cooperation in this case would entail higher tariffs as argued in Bagwell and Staiger (1990). However, computing cooperative tariffs requires assumptions about the bargaining procedure and is extremely computationally demanding, and hence will not be the focus of this paper.

[9] Existing works using temporary punishment usually emphasizes the role of renegotiation, which is not the focus of this paper. One example is Limao and Saggi (2008) which allows both the conventional infinite Nash reversion and a finite punishment phase in the repeated game. Quantifying renegotiation strategies usually involves additional

additional assumptions. Computing the minimum length of the punishment phase in the case of finite punishment, on the other hand, requires an estimate of the discount factor β as shown in the Appendix.

III Numerical Analysis

In this section, I first present a multi-country, multi-sector quantitative trade model. In this static model, governments can manipulate tariffs to improve welfare measured by either real expenditure or sectoral profits weighted by political incentives. I then use this model to quantify each country's static welfare gain when deviating from negotiated tariffs as well as subsequent welfare loss given means of punishment. The welfare changes in the punishment phase presented in Subsection III.2 include not only Nash tariffs and autarky but also two alternative scenarios that are less harsh. Lastly, I also compute each country's minimum patience and compares their propensity to deviate using the methodology developed in Section II.

III.1 A Static Trade Model

The economy consists of N countries and S industries. I use subscript i or j to index countries and s to index sectors. Each industry features monopolistic competition with constant elasticity of substitution. Homogeneous consumers have access to a continuum of differentiated varieties. The preference of a representative consumer in country j can be described by the following utility function:

$$U_j = \prod_s \Big[\sum_i \int_0^{M_{is}} x_{ijs}(\nu_{is})^{\frac{\sigma_s - 1}{\sigma_s}} d\nu_{is} \Big]^{\frac{\sigma_s}{\sigma_s - 1}\mu_{js}}, \tag{4}$$

where x_{ijs} is the quantity of an industry s product manufactured in country i consumed in country j, M_{is} is the mass of sector s varieties produced in country i, μ_{js} is the Cobb-Douglas share of country j's expenditure spent on sector s varieties, and $\sigma_s > 1$ is the elasticity of substitution across sector s varieties.

On the supply side, each variety is produced by exactly one firm. Firms within any sector are homogeneous and their technology is described by the following production function:

$$\sum_j \theta_{ijs} x_{ijs} = \zeta_{is} l_{is}, \tag{5}$$

where l_{is} is the labor an industry s firm requires in country i, θ_{ijs} represents iceberg trade barriers on

---

arbitrary assumptions and are much more computationally demanding.

sector s varieties from country i to country j, and ζ_{is} is the sectoral specific technology parameter. Note that M_{is} in (4) is exogenously determined. Hence, this model does not feature free entry, and firms are allowed to have positive profits[10]. In other words, we have $\pi_{is} \equiv M_{is} \sum_j (p_{is}\theta_{ijs}x_{ijs} - w_i l_{is})$, where π_{is} represents the sectoral profit for industry s in country i, p_{is} is the ex-factory price of sector s variety produced in country i, and w_i is the wage rate in country i.

Define Y_j as the country's nominal income and P_j as the aggregate ideal price index. Similarly, let Y_{js} be the nominal income of sector s in country j. Components of sectoral income Y_{js} is described by

$$Y_{js} = w_j L_{js} + \pi_{js} + \frac{L_{js}}{L_j} R_j,$$

where L_{js} is the employment of sector s in country j, L_j is exogenously given total employment in country j, and R_j is the total tariff revenue in country j. By definition, $Y_j = \sum_s Y_{js}$.

Tariffs are the only policy instrument allowed in this model. Let t_{ijs} be the ad valorem tariff country j imposes on country i's variety in sector s. To simplify notations, also define $\tau_{ijs} = 1 + t_{ijs}$. Government preferences are described by the following objective function:

$$W_j \equiv \sum_s \lambda_{js} W_{js}, \tag{6}$$

where $W_{js} \equiv \frac{Y_{js}}{P_j}$ is the real income of industry s in country j and $\lambda_{js} > 0$ is that industry's political weight. λ_{js} is scaled so that $\frac{1}{S}\sum_s \lambda_{js} = 1$. In other words, one dollar in industry s is equivalent to λ_{js} dollars from an industry with average political weight in the government's objective function. Therefore, in the benchmark case when the political weights are set equal to one, $W_j = \frac{Y_j}{P_j}$ and the government is simply maximizing the country's real income.

In equilibrium, households maximize their utilities subject to budget constraint given prices. Solving for the households' utility maximization gives the demand for varieties in sector s produced in country i:

$$x_{ijs} = \frac{(p_{is}\theta_{ijs}\tau_{ijs})^{-\sigma_s}}{P_{js}^{1-\sigma_s}} \mu_{js} Y_j, \tag{7}$$

where P_{js} is the ideal price index of sector s in country j defined by $P_{js} = \left[\sum_i M_{is}(p_{is}\theta_{ijs}\tau_{ijs})^{1-\sigma_s} \right]^{\frac{1}{1-\sigma_s}}$. Given (7), each firm from sector s in country i maximizes its profit by charging a constant markup

---

[10] The model can also be solved with free entry and fixed costs of production, as in Krugman (1980). I choose this version mostly because it rules out corner solutions with zeros productions in some sectors, so that the model can be solved numerically with a simpler algorithm.

over marginal costs:

$$p_{is} = \frac{\sigma_s}{\sigma_s - 1}\frac{w_i}{\zeta_{is}}. \tag{8}$$

The static equilibrium of this model can be characterized by the following system of equations:

$$\pi_{is} = \frac{1}{\sigma_s}\sum_j M_{is}\tau_{ijs}^{-\sigma_s}\left(\frac{\sigma_s}{\sigma_s - 1}\frac{\theta_{ijs}}{\zeta_{is}}\frac{w_i}{P_{js}}\right)^{1-\sigma_s}\mu_{js}Y_j \tag{9}$$

$$w_i L_i = \sum_s \pi_{is}(\sigma_s - 1) \tag{10}$$

$$Y_j = w_j L_j + \sum_i \sum_s t_{ijs} M_{is}\tau_{ijs}^{-\sigma_s}\left(\frac{\sigma_s}{\sigma_s - 1}\frac{\theta_{ijs}}{\zeta_{is}}\frac{w_i}{P_{js}}\right)^{1-\sigma_s}\mu_{js}Y_j + \sum_s \pi_{js} \tag{11}$$

$$P_{js} = \left[\sum_i M_{is}\left(\frac{\sigma_s}{\sigma_s - 1}\frac{w_i\theta_{ijs}\tau_{ijs}}{\zeta_{is}}\right)^{1-\sigma_s}\right]^{\frac{1}{1-\sigma_s}}. \tag{12}$$

The first two conditions are derived by substituting (5), (7), and (8) into the definition of sectoral profits and the labor market clearing condition $L_i = \sum_s M_{is} l_{is}$, respectively. (11) involves substituting demand for x_{ijs} and the firm's optimal pricing into the budget constraint. The last condition is obtained by simply replacing p_{is} in the formula of the sectoral ideal price index by (8). The problem of solving this system directly is that unknown parameters $\{M_{is}, \theta_{ijs}, \zeta_{is}\}$ are difficult to estimate empirically.

To circumvent this problem, I adopt the method popularized by Dekle et al. (2007). In particular, define Γ_{ijs} as the factual value of sector s trade flowing from country i to country j evaluated at world prices:

$$\Gamma_{ijs} \equiv M_{is}\left(\frac{\sigma_s}{\sigma_s - 1}\frac{\theta_{ijs}}{\zeta_{is}}\frac{w_i}{P_{js}}\right)^{1-\sigma_s}\tau_{ijs}^{-\sigma_s}\mu_{js}Y_j.$$

Then the conditions described by (9)-(12) can be rewritten in changes:

$$\hat{\pi}_{is} = \sum_j \alpha_{ijs}(\hat{\tau}_{ijs})^{-\sigma_s}\left(\frac{\hat{w}_i}{\hat{P}_{js}}\right)^{1-\sigma_s}\hat{Y}_j \tag{13}$$

$$\hat{w}_i = \sum_s \delta_{is}\hat{\pi}_{is} \tag{14}$$

$$\hat{Y}_j = \frac{w_j L_j}{Y_j}\hat{w}_j + \sum_i \sum_s \frac{t_{ijs}\Gamma_{ijs}}{Y_j}\hat{t}_{ijs}(\hat{\tau}_{ijs})^{-\sigma_s}\left(\frac{\hat{w}_i}{\hat{P}_{js}}\right)^{1-\sigma_s}\hat{Y}_j + \sum_s \frac{\pi_{js}}{Y_j}\hat{\pi}_{js} \tag{15}$$

$$\hat{P}_{js} = \left[\sum_i \gamma_{ijs}(\hat{w}_i\hat{\tau}_{ijs})^{1-\sigma_s}\right]^{\frac{1}{1-\sigma_s}}, \tag{16}$$

where the "hat" variables denote the ratios between the counterfactual and factual values, $\alpha_{ijs} \equiv \frac{\Gamma_{ijs}}{\sum_n \Gamma_{ins}}$, $\gamma_{ijs} \equiv \frac{\tau_{ijs}\Gamma_{ijs}}{\sum_s \tau_{mjs}\Gamma_{mjs}}$, and $\delta_{is} \equiv (\sum_j \frac{\sigma_s - 1}{\sigma_s}\Gamma_{ijs})/(\sum_t \sum_n \frac{\sigma_t - 1}{\sigma_t}\Gamma_{int})$.

Equations (13)-(16) represent a system of $2N(S+1)$ equations with $2N(S+1)$ unknowns $\{\hat{w}_i, \hat{Y}_i, \hat{P}_{is}, \hat{\pi}_{is}\}$. Compared to the system represented by (9)-(12), this system has the following advantages: first, the coefficients depend only on σ_s, λ_{js}, and observables so that information on $\{M_{is}, \theta_{ijs}, \zeta_{is}\}$ is no longer needed. In addition, all observables can be inferred directly from widely available trade and tariff data because $Y_j = \sum_i \sum_s \tau_{ijs}\Gamma_{ijs}$ and $w_j L_j = Y_j - \sum_i \sum_s t_{ijs}\Gamma_{ijs}$, where $\pi_{is} = \frac{1}{\sigma_s}\sum_j \Gamma_{ijs}$. Table 1 lists σ_s and averages of t_{ijs} and λ_{js} for all 33 sectors. More discussions of data and calibration can be found in Section A of the Appendix.

Given any counterfactual tariffs, we can solve equations (13)-(16) simultaneously and then calculate welfare changes relative to the factual equilibrium by

$$\hat{W}_j = \frac{\sum_s \lambda_{js} Y'_{js}}{\hat{P}_j \sum_s \lambda_{js} Y_{js}},$$

where Y'_{js} is the counter-factual income of sector s in country j. Note that the presence of aggregate trade imbalances in the data is not coherent with this model. Hence, I follow the exercise in Dekle et al. (2007) to construct a trade flow matrix without trade imbalance. All later calculations of welfare changes given counterfactual tariffs will treat this purged trade flow data as the factual equilibrium.

III.2 Welfare Changes in Stage Game

Table 2 presents each country's welfare changes in the deviation phase. Entries under "Own" are maximum one-period percentage gain in welfare when one country deviates but other countries still impose tariffs in the static cooperative equilibrium, calculated following (1). Entries under "Other," on the other hand, are averages of other countries' one-period welfare loss in the deviation phase. Computing these welfare changes involves maximizing the government's objective function (6) subject to equilibrium conditions (13)-(16). The first two columns show welfare changes when political weights are set to be one for all sectors in all countries. The last two columns take political factors into consideration by including the calibrated political weight parameter λ_{js} into the government's objective function.

From Table 2, we can see that all countries can gain at the expense of other countries in the deviation stage. Therefore, following the equilibrium strategy discussed in Section II, any country's deviation will trigger a punishment phase in the next period. Also note that incorporating political weights into governments' objective functions do not change the results substantially. Comparing to the case without political weights, the maximum welfare gain in the deviation phase increases for four countries (South Korea, the United States, European Union, and Rest of World), but deceases for the other six.

Table 3 and table 4 present one-period percentage welfare changes of the deviating country in the punishment phase, both with and without political weights. Both tables consider four scenarios in the punishment phase. Among them, "Trade War" and "Autarky" are commonly used as punishment in theoretical analysis that models trade policy cooperation as a repeated game. In addition, I also consider two scenarios that involve less severe punishment and are hence more realistic. "WTO Bound" refers to the static equilibrium in which victim countries impose the bound rates (the maximum rates allowed by the WTO) against the deviating country, whereas the deviating country imposes optimal tariffs in response. On the other hand, "Sanction" refers to the static equilibrium in which all other countries sanction the deviating country by imposing optimal tariffs against the deviating country only, but sustain negotiated tariffs with each other.[11] Note that for all four scenarios in the punishment phase, retaliatory tariffs span all 33 agricultural and manufacturing industries. This is actually consistent with the DSU rules articulated in WTO (1994): whereas WTO only allows retaliatory tariffs to be imposed in the same sector as that in which the violation was found (Article 22.3(a) of the DSU), it is also specified in Article 22.3(f) of the DSU that all goods belong to the same sector. For example, when a country imposes high tariffs on automobiles, other countries can counter with retaliatory tariffs on agricultural products.

Computing welfare changes under "WTO Bound" is done by first substituting WTO bound rates as victim countries' counter-factual tariffs, and then solving for the deviating country's optimal tariffs given other countries' punishment tariffs. Computing entries under "Sanction" and "Trade War" both involve iterating the algorithm used to calculate optimal tariffs and re-optimizing until the best-response equilibrium is found. The difference is that in the sanction case, each country can only optimize over tariffs against the deviating country, whereas in the trade-war case, tariffs against all other countries are allowed to change. Similar to "WTO Bound", the deviating country imposes optimal tariffs given other countries' punishment tariffs in these two scenarios. Lastly, welfare changes in the case of autarky can be obtained by solving (13)-(16) numerically with very high counterfactual tariffs.[12]

From Table 3 we can see that imposing WTO-bound tariffs is the weakest form of punishment compared to the other three scenarios: no deviating country experiences a welfare loss in the punishment phase in the case without political weights. In other words, if other countries react to a deviation by imposing WTO bound rates in the punishment phase, the cooperative static

---

[11] Among the four scenarios in the punishment phase, Trade War and Autarky are Nash equilibrium in the stage game but the other two are not.

[12] Computations that involve optimization are repeated multiple times with different initial values. Sometimes the resulting tariffs are different, but welfare changes are stable because welfare is very flat in tariffs in the neighborhood of its maximum.

equilibrium is not sustainable. One interpretation of this result is that when one country abandons the WTO rules and impose high tariffs, punishments that still follow WTO rules are no longer considered effective deterrents.

All other three scenarios presented in Table 3 are harsher punishments than imposing bound rates as shown in the average welfare loss. In these three punishment-phase scenarios, the only case in which the static cooperative equilibrium is unsustainable is Japan under sanction. The positive welfare change under sanction is mainly due to Japan's extreme factual trade policy in agricultural industries.[13] Under sanction, although other countries impose punitive tariffs against Japan, our definition of the scenario allows Japan to optimize over its tariffs. As a result, the country is able to abolish the existing high factual tariffs, which actually leads to welfare improvement.

The last row of Table 3 shows the magnitude of the average welfare loss in trade war is larger than that in sanction. This result is in line with the idea of the punishment effect discussed in Bagwell and Staiger (1999b). In the punishment phase, imposing high tariffs against the deviating country will increase trade volume among non-deviating countries. This trade diversion does not generate much tariff revenue under sanction, because tariffs among non-deviating countries are fixed at low levels. To the contrary, high tariffs among non-deviating countries under trade war imply higher tariff revenue from trade diversion, creating additional incentives to impose higher tariffs against the deviating country. Evidence that corroborates this explanation is presented in Table 5, which lists the factual median tariffs and median tariffs in sanction and trade war with no political weights. Under "Own" are median tariffs a country imposes against other countries, whereas under "Other" are median tariffs imposed by other countries against that specific country. We can see that, "Own" median tariffs for all countries do not differ significantly between sanction and trade war, whereas "Other" tariffs are much lower in the sanction scenario.

Table 4 present one-period percentage welfare changes of the deviating country in the punishment phase when political weights are incorporated. Overall, the patterns observed in the benchmark case presented in Table 3 still hold when political weights are taken into consideration. One difference is that now India and Mercosur's welfare change in the case of "WTO Bound" becomes negative. This is because for these two countries, the sectors with high political weights also face high bound tariffs. For example, both countries have highest political weight in the sector of wheat (1.67 for India and 1.34 for Argentina). Meanwhile, the average bound tariff imposed on wheat is also very high for these two countries (45.5% and 48.6%, respectively).

---

[13] As can be seen in Figure 1, the distribution of Japan's tariff across sectors is very skewed. Japan's factual tariff on rice, wheat, and dairy are up to 513%, 73%, and 54%, respectively. On the other hand, its median tariff is only 1%.

III.3 Minimum Patience

Given Ω_j and Φ_j, I can then compute the minimum discount factor needed to sustain the static cooperative equilibrium $\underline{\beta_j}$ using (3). Table 6 presents the resulting $\underline{\beta_j}$ in the benchmark case for all scenarios in the punishment phase, whereas Table 7 presents the results when political weights are incorporated. For all four scenarios, the average minimum patience is smaller when political weights are included. However, the magnitude of difference is not very large and the patterns of the two tables are very similar.

For both tables, the computed minimum patience needed to sustain the static cooperative equilibrium exhibits a descending pattern from "WTO Bound" in the first column to "Autarky" in the last. This pattern reflects the strength of the punishment strategies and hence the ability to deter deviation. Under "WTO Bound" which is the weakest means of punishment, all countries will deviate from the static cooperative equilibrium in the benchmark case. In other words, when one country has incentive to abandon the WTO rules and deviate from negotiated low tariffs, punishments that are constrained by WTO are no longer effective deterrents. On the other hand, the computed $\underline{\beta_j}$ is much smaller than one for all countries under "Trade War" and "Autarky". Consistent with most existing theoretical findings using the repeated game framework, these two means of punishment can effectively deter deviations from low cooperative tariffs.

The computed minimum patience $\underline{\beta_j}$ in Table 6 and Table 7 can also be used to compare the propensity to deviate across countries. For example, Japan is always among the most likely to deviate for any given method of punishment. In fact, it is the only country that will deviate if the punishment strategy involves infinite reversion to sanction. This is because Japan's gain from deviation is large whereas the loss in the punishment phase is mild relative to other countries. On the other hand, Russia and Mercosur are in general less likely to deviate compared to other countries.

To summarize, I first develop a static trade model in Subsection III.1, and then use it to quantitatively characterize countries' equilibrium strategies in a repeated-game framework. The procedure consists of two steps. First, I compute countries' static welfare changes in the deviation phase and the punishment phase. Next, I calculate the minimum patience (discount factor) needed to satisfy the self-enforcing constraint of each country. I also use this measure to distinguish whether the static cooperative equilibrium is sustainable under different scenarios of the punishment phase and compare countries' propensity to deviate from factual tariffs. From the results presented above, the effect of incorporating the political weights into the analysis is limited. In the following sections, I will use the procedure developed in this section to quantitatively analyze some theoretical findings in the trade-policy literature.

IV Trade Cooperation and Multilateral Institutions

This section is devoted to using the methodology discussed in Section II to analyze the role of the multilateral enforcement mechanism in trade cooperation. In Maggi (1999), the multilateral enforcement mechanism is defined as a punishment strategy whereby retaliative tariffs from all countries follow any deviation. This contrasts to bilateral enforcement mechanism, in which only the victim country retaliates. The author constructs a simple partial equilibrium trade model in a repeated game, and demonstrates that "in the presence of bilateral imbalances of power, countries can sustain a higher symmetric equilibrium payoff with multilateral enforcement than with bilateral enforcement." The rationale behind this result is straightforward: under the bilateral enforcement mechanism, a country will impose high tariffs against its weakest trading partners, because the strength of retaliation will be limited. On the other hand, under multilateral enforcement mechanism, punitive tariffs can also be imposed by third parties. This mechanism increases loss from deviation and hence reduces maximum sustainable cooperative tariffs, which in turn result in higher sustainable equilibrium payoffs.

In this section, I analyze the role of multilateral institutions quantitatively with a similar repeated-game framework studied in Maggi (1999), but with the more comprehensive general equilibrium trade model constructed in Section III. Welfare changes in the deviation and punishment phase as well as the minimum patience the under the bilateral enforcement mechanism are computed and compared with those under the multilateral enforcement mechanism. Such comparison allows us to quantify the impact of multilateral institutions on countries' equilibrium strategies and the sustainability of trade-policy cooperation. The numerical results corroborate Maggi's argument qualitatively: factual tariffs will not be sustainable in a world without a multilateral enforcement mechanism, and countries will impose deviating tariffs against its relatively weak trading partners. However, in the new equilibrium with only a bilateral enforcement mechanism, the magnitude of welfare changes is rather small.

Note that although not stated explicitly, the analysis presented in Section III actually relies on the following two assumptions:

1. All other countries perfectly observe any country's deviation from cooperative tariffs.
2. Once a deviation from negotiated tariffs is observed and some country experiences a welfare loss, all other countries are obligated to participate in the punishment phase, regardless of whether their welfare is compromised.

Given these assumptions, even if a country imposes optimal tariffs against only one country, all other countries will retaliate. Hence, when any country deviates, it prefers to impose optimal tariffs against all other countries. In a way, my analysis assumes the existence of some international organization such as the WTO to inform the trading community and execute punishment when a deviation from negotiated tariffs occurs. These implicit assumptions resemble the idea of the multilateral enforcement mechanism discussed in Maggi (1999).

To examine whether this mechanism improves the sustainability of trade cooperation, I apply the same procedure developed in Section II, but drop the two assumptions stated above. In other words, now no country observes the deviation or is allowed to retaliate, unless it is the victim country in the deviation phase. As a result, not imposing optimal tariffs against all countries in the deviation phase may be rational for the deviating country. For illustration purposes, I will focus on sequential equilibria in which countries can impose optimal tariffs only to one other country in the deviation phase, but the same analysis can be applied to the case of deviating against multiple countries. Under bilateral enforcement, each country has nine potential targets, so we need to analyze 90 possible deviations instead of just 10 under multilateral enforcement. For each deviation, country j chooses tariffs only against country i to maximize its objective function (6) subject to equilibrium constraints (13)-(16), while keeping its tariffs against other countries unchanged at previously negotiated levels. All the quantitative results presented in this section are computed by excluding the political weights in the government's objective function.[14]

Table 8 summarizes welfare changes in the deviation phase. The first column lists the deviating countries, and the entries under "Self," "Target," and "Others" are summary statistics of one-period welfare changes of the deviating country, the victim country, and other countries, respectively. Compared to the values in Table 2, we can see the magnitude of welfare changes under country-specific deviation is smaller in general. This result is expected, because the deviating country now can only manipulate tariffs against one other country. In addition, the results also indicate considerable variation in gains from deviation. For example, Canada's welfare gain can be as large as 1.6% (targeting the United States) or as low as practically nothing (targeting Russia). Clearly, such heterogeneity is related to the volume of factual trade between the deviating country and the target country. Significant variation is also observed in the target country's welfare loss. In general, deviation from large economies (e.g., the United States and EU) can reduce the welfare of their target countries to a greater extent. One surprising case is that when Korea imposes optimal tariffs against India in the deviation phase, India's welfare actually increases by 0.01%. This result is due to the distorting factual tariffs imposed by Korea. For example, its tariff on the oil seed industry imports from India is 612%, which is the largest factual tariff in our data set. But the calculated optimal tariff on the oil seed industry when Korea deviates against India is 53%. The fact that the calculated optimal tariffs in some industries are actually lower than factual tariffs leads to a slight welfare improvement of the target country.

---

[14] As a robustness check, I also compute the corresponding welfare changes when the political weights are included. The main arguments presented in this section still hold qualitatively.

After calculating welfare gains in the deviation phase, the next step is to quantify corresponding static welfare changes in the punishment phase. Following the analysis in Maggi (1999), I focus on the scenario in which Nash tariffs are used in the punishment phase. Under bilateral enforcement, the trade war equilibrium in the stage game refers to the case in which both the deviating country and the target country impose Nash tariffs against each other while keeping tariffs against other countries unchanged. To calculate such welfare changes in a bilateral trade war equilibrium, I simply iterate the algorithm used to calculate results in Table 8 for the two involved countries and re-optimize until a best-response equilibrium is found.

Figure 2 presents the static welfare changes of the trade war case. Because which country deviates and hence the order of country pair does not matter in bilateral trade war, Figure 2 is set up in such a way so that the country with greater welfare change is always in the x-axis. In the figure, we observe three categories of welfare changes. The most abundant blue dots represent country pairs with negative welfare changes for both countries. This is the case in the punishment phase on which most studies in trade-negotiation literature focus. Any equilibrium represented by a blue dot is sustainable (given that the discount factor is not too small), because the welfare loss in the punishment phase will deter the potential deviating country. Also, the several green dots indicate one country in the country pair has a positive welfare change. This is the scenario first described in Johnson (1954), in which the bilateral trade war has a winner. If the winner happens to be the deviating country, cooperation between these two countries is not sustainable. Lastly, the one dot in red represents bilateral trade war between Korea and India, and the welfare change is positive for both countries. As explained previously, the high distorting factual tariffs Korea imposes is the reason for this abnormal result. Obviously, the negotiated tariffs between these two countries are not sustainable under bilateral enforcement mechanism. Comparing Figure 2 and Table 2, we can see the rationale presented in Maggi (1999) also works in the more comprehensive general equilibrium model in this paper: factual tariffs can be sustained under multilateral enforcement but not under bilateral enforcement when Nash tariffs are used as punitive tariffs in the punishment phase.

Table 9 presents some country-pair specific features of the 11 cases in which the bilateral trade war has a winner (excluding the Korea-India case). Two of the three features display different patterns between the winners and the losers. The first one is average tariffs: winning countries on average impose higher factual tariffs on imports from the losing countries. Japan and India impose relatively high tariffs, and both countries are winners in three bilateral trade wars. In addition, the winning countries on average also import more (as share of total imports) from the losing countries than vice versa. On the other hand, the average export share does not appear different substantially between the winning and losing countries.

The welfare consequences of removing the multilateral enforcement mechanism are summarized under “Welfare” in Table 10. Entries under “ΔW” are the welfare changes in the new counterfactual equilibrium. In this equilibrium, all 12 pair of countries with unsustainable factual tariffs impose bilateral Nash tariffs against each other, but keep factual tariffs with other countries. The following two columns are the number of cases in which a country will deviate and be deviated against, respectively. For example, China will deviate against two other countries, be deviated against by one other country, and its welfare change in the new equilibrium is 0.03%. In general, welfare change is positively related to the number of pairs in which the country chooses to deviate, and negatively related to the number of pairs in which the country is deviated against. Overall the welfare changes are small, and some countries gain at the expense of others. Figure 2 to a large extent preludes this result: the total number of red and green dots is significantly less that that of blue dots. In addition, the welfare changes of red and green dots are mostly small. Therefore, although the quantitative results presented in this section corroborate the mechanism described in Maggi (1999), the magnitude of welfare consequences from removing multilateral institutions is rather insignificant.

I also use (3) to calculate the minimum patience needed to sustain factual negotiated tariffs (which will be denoted by $\underline{\beta_j}^{BI}$) for the country pairs that are still sustainable under the bilateral enforcement mechanism. Comparisons between $\underline{\beta_j}^{BI}$ and $\underline{\beta_j}^{M}$, the minimum patience under the multilateral enforcement mechanism, are presented under “Sustainability” in Table 10. We can see that propensity to deviate only increases in less than half the remaining country pairs. For those country pairs that remain sustainable under bilateral enforcement, removing multilateral institutions does not appear to significantly reduce sustainability of trade cooperation.

Lastly, I explore the sustainability of the new equilibrium in which winners of the bilateral trade war deviate under the bilateral enforcement mechanism. In particular, I assume that the deviating countries and their corresponding victims countries impose bilateral Nash tariffs against each other whereas other countries keep imposing cooperative tariffs. Treating this stage equilibrium in the punishment phase as the new “cooperative” tariffs, I then compute the static welfare changes in the second round of the punishment phase in which both the deviating country and the target country impose Nash tariffs against each other. Comparing to the initial stage equilibrium with 11 country pairs imposing Nash tariffs, now two more pairs will have a winner of the bilateral trade war and the new “cooperative equilibrium” is still not sustainable under bilateral enforcement.

V Impact of Free Trade Agreements

In this section, I use the model developed in Section III to quantitatively analyze the impact of FTAs on countries’ equilibrium strategies and hence sustainability of trade cooperation. I first quantify the two major channels through which FTAs can affect trade-policy cooperation: the tariff complementarity effect and punishment effect. These two effects have been studied from a theoretical perspective, but I am not aware of any existing research that tries to quantitatively identify them. To preview the main results of this section, my analysis on bilateral FTAs shows the tariff complementarity effect reduces average median tariffs in the deviation phase by 12.7%, whereas the punishment effect in the trade war scenario reduces average median punitive tariffs in the punishment phase by 11.6%. In addition, the average propensity to deviate for all 10 countries increases if two other countries negotiate an FTA. However, for a member country, the post-FTA propensity to deviate can either increase or decrease.

Whereas almost all existing studies on how FTAs influence trade-policy cooperation use a repeated-game framework, I find the analysis in Bagwell and Staiger (1999b) particularly helpful in understanding the quantitative results presented in this section. Assuming countries can deviate from negotiated factual tariffs but not from any FTA, the authors construct an export-competing, partial equilibrium model and discuss the possible effects of forming FTAs on sustaining trade cooperation, two of which are relevant to this paper. The first effect is the tariff complementarity effect: when two countries form an FTA, they have less incentive impose high external tariffs. In other words, let t_{ij}^{DF} and t_{ij}^{D} denote country j's optimal deviation tariffs against country i before and after entering an FTA with some country other than i. Then the tariff complementarity effect is equivalent to $t_{ij}^{DF} < t_{ij}^{D}$. The rationale behind this effect is as follows: raising tariffs against country i will divert trade flows to other countries. Compared to the case without FTAs, zero tariffs between country j and its FTA partner means less tariff revenue generated from diverted trade flows, which leads to less incentive for country j to impose high tariffs against country i in the deviation phase. In the framework presented in this paper, this effect implies $\Omega_j(t_{ij}^{DF}) < \Omega_j(t_{ij}^{D})$. Therefore, via the tariff complementarity effect, signing new FTAs will decrease member countries' propensity to deviate, thus enhancing the sustainability of trade cooperation.

The other effect discussed in Bagwell and Staiger (1999b) is the punishment effect, which can be thought of as the flip side of the tariff complementarity effect in the punishment phase. This effect can be illustrated by the following example: country i and country j have the opportunity to enter a FTA, and country k is the external trading partner. When country k deviates, its welfare loss in the punishment phase depends on punishment tariffs imposed by country i and country j. If country i and country j negotiate a FTA, they have less incentive to impose high tariffs against country k in the punishment phase, just like the tariff complementarity effect in the deviation phase. Hence, the external country will face less future welfare loss after deviation. In this paper, the punishment effect is equivalent to a smaller $|\Phi_k|$. This effect should increase country k's propensity to deviate, which reduces sustainability of trade cooperation.

To examine whether these two effects exist in the more comprehensive general equilibrium trade model in this paper, I first compute all 45 counterfactual equilibria in which two of the 10 countries negotiate a FTA and impose zero tariffs on each other.[15] Treating each counterfactual equilibrium as the cooperative static equilibrium, I then calculate optimal tariffs and static welfare changes (for all 10 countries, regardless of whether they are FTA members) in the deviation phase. The trade-war scenario in punishment phase, because it is also the punishment strategy used in Bagwell and Staiger (1999b). Lastly, following the assumption in Bagwell and Staiger (1999b), deviation from the FTA is not allowed in either the deviation or the punishment phase.

Table 11 summarizes the main results related to the deviation phase. Entries under "FTA" are average percentage welfare changes and average median optimal tariffs of each country in the deviation phase when the country is also a FTA member in the counterfactual equilibrium (nine equilibria are possible for each country). Entries under "non-FTA" are corresponding values if the country is not in the FTA (36 equilibria in total for each country). I also list welfare changes and median optimal tariffs calculated from factual data under "Factual" for comparison purpose. All welfare changes are relative to the welfare levels in the actual equilibrium instead of counterfactual FTA equilibrium. Compared to entries under "Factual" in Table 11, mean welfare changes and mean median tariffs under "non-FTA" are almost identical. On the other hand, mean median tariffs under "FTA" are lower than those under "Factual" for all 10 countries, indicating the presence of the tariff complementarity effect. On average, median optimal tariffs of FTA member countries is 12.7% less than median optimal tariffs calculated from factual data. In addition, a country's welfare gain in the deviation phase after negotiating an FTA with another country on average decreases by 0.5%, or 20% of the welfare gain before entering the FTA. Note this significant shrink in welfare gain results not only from the tariff complementarity effect, but also from the fact that now the deviating country cannot impose optimal tariffs on its FTA partner.

Table 12 summarizes the main results under the trade war scenario in the punishment phase. Note that whereas entries under "FTA" and "nonFTA" are averages of 9 and 36 equilibria respectively, those entries under "Factual" are welfare changes and median tariffs of one trade war equilibrium. We can see that all 10 countries' welfare loss in the punishment phase is less than that calculated with factual data, independent of whether the country is a FTA member. However, explanations for such changes do depend on whether the country is a member. For the non-member countries, the decrease in welfare loss in a trade war is due to the punishment effect as discussed in Bagwell and Staiger (1999b): the average median tariffs imposed by FTA countries, displayed under the fourth column in Table 12, are on average 11.6% less than those under "Factual." This effect is unnoticeable for FTA countries, because the average median tariffs from non-FTA countries are almost the same as those under "Factual." Less welfare loss of the FTA member countries is attributable to enforceable FTA contracts: their FTA partners can only impose zero tariffs against

---

[15] For illustration purposes I only consider the case of bilateral FTA. Moreover, all numerical results in this section are computed using the benchmark model in which political weights are not incorporated.

them even in a trade war.

Another pattern shown in Table 12 is that all countries' welfare loss as FTA members is less than the loss outside FTA. In other words, the magnitude of the welfare change for a non-FTA country due to the punishment effect from two FTA countries is on average less than the magnitude of the welfare change for the same country due to zero tariffs from its FTA partner. This result implies a country's welfare is convex in other countries' tariffs, which is consistent with existing theory.

Finally, I use equation (3) to calculate the minimum patience needed to sustain the equilibrium in which two countries negotiate a FTA. The results are presented in Table 13. In all cases, the calculated $\underline{\beta_j}$ is less than one. Comparing the last two columns, we can see the average $\underline{\beta_j}$ for all countries under "non-FTA" are larger than those under "Factual." This result is consistent with the results shown in Table 11 and Table 12: when two other countries negotiate a FTA, a non-member country's gain from deviation does not change, but its loss in trade war decreases due to the punishment effect. Therefore, new FTAs actually make outside countries more likely to deviate from negotiated current tariffs. On the other hand, no such pattern is found for entries under "FTA." Compared to "Factual," India and Canada's average $\underline{\beta_j}$ increases while that for the other eight countries decreases, because both the FTA member countries' welfare gain from deviation and welfare loss in trade war decrease. Depending on the relative magnitude of changes in gain and loss, the propensity to deviate can change either way.

VI Sustainability of Future Trade Liberalization

This section utilizes the procedure developed in Section II to analyze how equilibrium strategies and hence sustainability of negotiated tariffs vary with exogenous multilateral trade liberalization. Contrary to Section V in which trade liberalization is bilateral, I now focus on the counterfactual scenario that trade liberalization takes place globally and symmetrically. In other words, assuming every country reduces tariffs to ρ percent of the original level, we can then treat $(\rho t_j^C, \rho t_{-j}^C)$ in the cooperative static equilibrium as factual tariffs, and quantify equilibrium strategies in these "counterfactual cooperative equilibria." In this setup, ρ can be considered a measure of trade liberalization from 1 (factual level of tariffs) to 0 (free trade). Therefore, in the remainder of this section, ρ will also be used to refer to the counterfactual cooperative equilibria that represent different stages of global trade liberalization . The main results are exhibited in Figure 3, which plots $\underline{\beta_j}$ for all countries with given ρ using trade war as the scenario in the punishment phase. The plot shows the counterfactual cooperative equilibrium is sustainable regardless of the value of ρ. In addition, with the exception of Japan, all other countries' propensity to deviate does not vary much with the extent of trade liberalization.

The setup explored in this section is related to Bagwell and Staiger (1990), in which the authors analyze the impact of trade volume on trade-policy cooperation. By analyzing a repeated game that incorporates a two-country partial equilibrium model as a stage game, they show the level of protection must rise in a cooperative equilibrium in periods of high trade volume to hold the incentive to deviate in check. Testing this theoretical result directly is possible by assuming exogenous changes in trade volume and calculating how the lowest sustainable cooperative tariffs vary accordingly. However, calculating the cooperative tariffs of 10 countries and 33 sectors is extremely computationally demanding and requires further assumption on the bargaining procedure (see Ossa, 2014). In addition, predicting changes in trade volume in future trade liberalization is difficult to model and involves further arbitrary assumptions. My approach that assumes symmetric reduction in tariffs requires far less computational power and resonates with the principle of reciprocity, which is embodied in the WTO practice.

To plot Figure 3, I first compute 20 different equilibria with symmetric tariff reduction ranging from 5% to 100% (free trade). For each ρ, I then compute each country's welfare change in both the deviation phase and global trade war in the punishment phase,treating $(\rho t_j^C, \rho t_{-j}^C)$ as the static cooperative equilibrium.[16] Notice that in Bagwell and Staiger (1990), gain from deviation depends on trade volume but expected loss in the punishment phase is fixed. Under my approach, however, both gain and loss will change with ρ.

When analyzing welfare changes in equilibria with various degrees of trade liberalization, welfare changes relative to the factual static cooperative equilibrium (t_j^C, t_{-j}^C) are different from those relative to the counterfactual cooperative equilibria $(\rho t_j^C, \rho t_{-j}^C)$. Let $\Omega_j^F(\rho)$ denote country j's maximum welfare gain when deviating from counterfactual equilibrium ρ, measured relative to its welfare in the factual equilibrium, and let $\Psi_j(\rho)$ denote the country j's welfare change from the factual equilibrium to the counterfactual equilibrium ρ. Then the deviation gain in equilibrium ρ relative to welfare in the counterfactual cooperative equilibrium, denoted by $\Omega_j^{CF}(\rho)$, can be calculated by the following equation:

$$\Omega_j^{CF}(\rho) + 1 = \frac{\Omega_j^F(\rho) + 1}{\Psi_j(\rho) + 1}. \tag{17}$$

$\Omega_j^{CF}(\rho)$ instead of $\Omega_j^F(\rho)$ is the gain from deviation that should be used in calculating the minimum length of the punishment phase in counterfactual equilibrium ρ. In theory, as ρ increases, aggregate welfare should also improve in the counterfactual equilibrium due to increasing trade volume. Previous research on trade-policy theory also predicts that, relative to the factual equilib-

---

[16] Trade war is also the punishment strategy considered in Bagwell and Staiger (1990). In addition, sanction is not considered because of the computational constraint: for every value of ρ, 10 possible sanction scenarios need to be computed as compared to just one trade war scenario.

rium, average welfare gain from deviation is larger when the tariff level in the cooperative stage game is lower. Because both the numerator and the denominator of the right-hand side of (17) are increasing in ρ, no prediction on how $\Omega_j^{CF}(t_j^D, \rho)$ changes in ρ can be formed from theory. Figure 4 plots the two calculated average welfare gains from deviation. As expected, the average gain from deviation relative to factual equilibrium increases monotonically with trade liberalization. A similar pattern is also found for that relative to counterfactual equilibrium. Combining this result with (17), we can see the magnitude of the average increase in Ω_j^F is larger relative to that in $\Psi_j(\rho)$.

The relationship between welfare changes in the punishment phase and trade liberalization can be analyzed analogously. Using the same notations, we have the following:

$$\Phi_j^{CF}(\rho) + 1 = \frac{\Phi_j^F(\rho) + 1}{\Psi_j(\rho) + 1} \tag{18}$$

where $\Phi_j^{CF}(\rho)$ and $\Phi_j^F(\rho)$ are static welfare loss in the punishment phase relative to the ρ counterfactual equilibrium and the factual equilibrium, respectively. Because the scenario in the punishment phase I choose to focus on in this section is trade war, $\Phi_j^F(\rho)$ should not vary in ρ: the resulting Nash equilibrium in the stage game does not depend on tariff levels in counterfactual cooperative equilibrium. This prediction from theory is exactly what is presented in Figure 5, which plots static average welfare changes in the punishment phase. Because average welfare improves as all countries liberalize trade symmetrically, $\Phi_j^{CF}(\rho)$ also increases in magnitude along with trade liberalization.[17] In addition, the magnitude of changes in welfare loss is also comparable to that in welfare gain, as shown in Figure 4. Furthermore, calculations show $\Phi_j^{CF}(\rho)$ is negative in all counterfactual equilibria. Therefore, cooperation can always be sustained and exogenous shocks need to be introduced to analyze how countries' incentive to deviate varies with ρ.

Figure 6 illustrates how $\Psi_j(\rho)$ of each country varies with ρ. We can see that as ρ decreases from one to zero, welfare change varies significantly across countries. Japan appears to have the largest welfare gain of trade liberalization in static cooperative equilibrium, because moving from factual tariffs to free trade increases its welfare by almost 2%. Again, this result can be explained by Japan's high factual tariffs in certain industries. On the other hand, as the largest loser, Russia bears a welfare loss of about 0.7% in free trade. Also, note that welfare change is not always monotone in ρ. For example, India's welfare change increases until ρ moves to about 0.3% and then decreases when trade further liberalizes.

Figure 4 and Figure 5 jointly explain the pattern observed for all countries except Japan in Figure 3. As trade liberalizes in a multilateral manner, countries' average gain from deviation and

---

[17] From Figure 5, average welfare reaches its maximum when ρ is slightly greater than zero. This result is probably due to the average welfare not being weighted by country size. As Figure 6 shows, Russia's welfare decreases with trade liberalization, but its aggregate expenditure is small relative to other countries.

average loss in the punishment phase both increase. Hence, countries' propensity to deviate should not vary much with trade liberalization. The exceptional case of Japan is again due to its highly distortive factual tariffs. When ρ decreases, Japan's welfare gain by imposing optimal tariffs also decreases significantly. As shown in Table 3 and Table 4, Japan's welfare loss in trade war is the smallest among all countries. Therefore, even though its $\Psi_j(\rho)$ is also larger than other countries, the resulting $\Phi_j^{CF}(\rho)$ is still small relative to its $\Omega_j^{CF}(\rho)$. As a result, Japan's propensity to deviate reduces drastically along with worldwide trade liberalization.

VII Conclusion

In this paper, I propose a methodology to quantify countries' equilibrium strategies on tariffs in a repeated-game framework that has been widely used in the trade-policy literature. I characterize a country's equilibrium strategy by its gain from deviation, loss as a consequence of deviation, and the minimum patience such that the negotiated low tariffs are sustainable. To illustrate this methodology, I construct a reasonably comprehensive general equilibrium trade model and numerically compute the equilibrium strategies of major participants of international trade. Next, I use this model to study some existing theoretical findings that are yet to be quantitatively analyzed. First, I find that in a counterfactual scenario without a multilateral enforcement mechanism, factual tariffs are not sustainable. However, the welfare consequences are quantitatively insignificant. Then I move to analyze the impact of FTAs, and quantitatively investigate the tariff complementarity effect and the punishment effect. Comparing the propensity to deviate before and after FTAs are negotiated, I find that FTAs reduce the sustainability of multilateral cooperation in trade policy in general. Lastly, I show that, with the exception of Japan, other major countries' propensity to deviate does not vary much along with exogenous trade liberalization.

This paper is the first quantitative exercise of trade policy in a repeated game framework. By computing the welfare effects of trade policy in a reasonably comprehensive quantitative trade model, I can numerically examine the existing theoretical findings derived from simpler trade models. In addition to quantify the welfare effects of Nash tariffs and autarky which are commonly used in the theoretical analysis, I also incorporate two alternative means of punishments that are less stringent and hence more realistic. If the trade model used to quantify equilibrium strategies and calculate the propensity to deviate is taken to be realistic, the results presented can be treated as answers to questions of immediate policy relevance.

The methodology discussed in this paper can be applied to quantitatively analyze trade policy in many ways. For example, one can build a Melitz (2003) style model that features heterogeneous firms. Another interesting extension is to allow deviation and punishment to be industry specific, and to study the sustainability of cooperation at the industry level. Lastly, I abstract from collusion among countries throughout this paper. It may be interesting to investigate the scenario in which two or more countries deviate at the same time and maximize their welfare jointly.

Appendix

A Data

Similar to Ossa (2014), the trade, production, and applied tariff data come from the Global Trade Analysis Project (GTAP) 8 database. This database is based on a number of underlying databases. The database is documented in Narayanan, Aguiar and McDougall, eds (2012), which can be accessed directly from the GTAP website under http://www.gtap.agecon.purdue.edu. Bound tariff data comes from the International Trade Centre's Market Access Map database, which is also the original data source of GTAP's allied tariff data.

Demand elasticity estimation follows Ossa (2016), which uses the Feenstra (1994) method with the UN Comtrade data for the period 1994-2008. In addition, political weights for the United States, China, Japan, India and the European Union are taken directly from Ossa (2014). For the remaining five countries, I compute their political weights using the algorithm discussed in Ossa (2014). The main idea is to iterate the political weights until the optimal unilateral tariffs converge to the distribution of factual tariffs. Readers interested in more details can go to Section A, B, and D of the Appendix in Ossa (2014).

B Punishment Lasting Finite Number of Periods

Consider the following strategy of the repeated game:

1. Each country imposes cooperative tariffs t^C in every stage game unless some country deviated from cooperation in the past.

2. If some country has not played t^C in the past and some other countries incur a welfare loss in the deviation phase, every country will enter a punishment phase and impose t^P for K periods and then return to imposing cooperative tariffs.

Given discount factor β, one-period welfare gain from deviation $\Omega_j(t_j^D)$, and one-period welfare loss in the punishment phase $\Phi_j(t_j^P, t_{-j}^P)$, solving the following equation will give the minimum length of the punishment phase necessarily for country j to sustain cooperation:

$$\Omega_j(t_j^D) + \frac{\beta(1-\beta^K)}{1-\beta}\Phi_j(t_j^P, t_{-j}^P) = 0 \tag{B.1}$$

We can denote the solution to this equation as $\underline{K_j}$. Whereas the model features discrete time, I allow the resulting $\underline{K_j}$ to be a non-integer. Given that both Ω_j and Φ_j are positive and β is not too small, the calculated $\underline{K_j}$ will be a positive number. When the length of the punishment phase in the aforementioned strategy is larger than $\underline{K}_j$, it is a Nash equilibrium of the repeated game, in which the cooperative stage equilibrium is sustained.

Both infinite punishment and punishment with finite periods have been considered as equilibrium strategies in works that model trade policy cooperation in a repeated game, although infinite reversion (to Nash tariffs) is used more often. Both the minimum patience $\underline{\beta}_j$ and the minimum length of the punishment phase $\underline{K}_j$ can be used as measures of the sustainability of the cooperative equilibrium. For example, the cooperative equilibrium is unsustainable when $\underline{\beta}_j \geq 1$ or $\underline{K}_j < 0$. Even when the cooperative equilibrium is sustainable, countries are more likely to deviate when $\underline{\beta}_j$ increases or $\underline{K}_j$ decreases. In fact, the following lemma shows that comparing $\underline{\beta}_j$ between countries is equivalent to comparing $\underline{K}_j$ when analyzing the sustainability of the cooperative stage equilibrium.

Lemma 1 *Let $\underline{\beta}$ and $\underline{K}$ be the solution to* (3) *and* (B.1) *respectively. If country j and country l have the same discount factor β, then $\underline{\beta}_j > \underline{\beta}_l \leftrightarrow \underline{K}_j > \underline{K}_l$.*

Proof. From (3), $\underline{\beta}_j > \underline{\beta}_l$ implies $\frac{\Omega_j}{\Omega_j - \Phi_j} > \frac{\Omega_l}{\Omega_l - \Phi_l}$. Since $\Omega_j > \Phi_j$ and $\Omega_l > \Phi_l$, we have $\frac{-\Phi_j}{\Omega_j} < \frac{-\Phi_l}{\Omega_l}$. From (B.1), we have $\frac{\beta(1-\beta^{\underline{K}_j})}{1-\beta} = \frac{-\Omega_j}{\Phi_j}$ and $\frac{\beta(1-\beta^{\underline{K}_l})}{1-\beta} = \frac{-\Omega_l}{\Phi_l}$. Since $\beta \in (0,1)$, it must be the case that $\underline{K}_j > \underline{K}_l$. By the same procedure, we can also show that $\underline{K}_j > \underline{K}_l$ implies $\underline{\beta}_j > \underline{\beta}_l$. ■

Because of the equivalence of $\underline{\beta}_j$ and $\underline{K}_j$, either measure can be used as a measure of countries' propensity to deviate. When Nash tariffs or autarky is used in the punishment phase, the strategy involving infinite punishment phase is subgame perfect but the one involving finite punishment phase is not. In other words, in the case with finite periods of punishment, the punishing countries will not have any incentive to return to cooperative tariffs once the punishment phase commences. This difference does not affect the quantitative analysis in this paper.

References

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

_ and _ , "An Economic Theory of GATT," *American Economic Review*, 1999, *89* (1), 215–248.

_ and _ , "Regionalism and Multilateral Tariff Cooperation," in John Piggot and Alan Woodmand, eds., *International Trade Policy and the Pacific Rim*, London: MacMillan, 1999, pp. 157–85.

_ , Chad P. Bown, and Robert W. Staiger, "Is the WTO Passe?," *Journal of Economic Literature*, December 2016, *54* (4), 1125–1231.

Beshkar, Mostafa, "Trade skirmishes safeguards: A theory of the WTO dispute settlement process," *Journal of International Economics*, 2010, *82* (1), 35 – 48.

Bhagwati, Jagdish N., *The World Trading System at Risk*, Princeton University Press, 1991.

Bown, Chad P., "On the Economic Success of GATT/WTO Dispute Settlement," *The Review of Economics and Statistics*, 2004, *86* (3), 811–823.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Unbalanced Trade," *American Economic Review*, 2007, *97* (2), 351–355.

Dixit, Avinash, "Strategic aspects of trade policy," in Thruman F. Bewley, ed., *Advances in Economic Theory: Fifth World Congress*, New York, NY: Cambridge University Press, 1987, pp. 329–62.

Eaton, Jonathan and Samuel Kortum, "Technology, Geography, and Trade," *Econometrica*, 2002, *70* (5), 1741–1779.

Estevadeordal, Antoni, Caroline Freund, and Emanuel Ornelas, "Does Regionalism Affect Trade Liberalization Toward Nonmembers?," *The Quarterly Journal of Economics*, 2008, *123* (4), 1531–1575.

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *The American Economic Review*, 1994, *84* (1), 157–177.

Grossman, Gene M. and Elhanan Helpman, "Protection for Sale," *The American Economic Review*, 1994, *84* (4), 833–850.

Horn, Henrik, Louise Johannesson, and Petros C. Mavroidis, "The WTO Dispute Settlement System 1995-2010: Some Descriptive Statistics," Working Paper Series 891, Research Institute of Industrial Economics December 2011.

Johnson, Harry G., "Optimum Tariffs and Retaliation," *Review of Economic Studies*, 1954, *21* (2), 142–153.

Krugman, Paul, "Scale Economies, Product Differentiation, and the Pattern of Trade," *American Economic Review*, 1980, *70* (5), 950–959.

Limao, Nuno, "Preferential Trade Agreements as Stumbling Blocks for Multilateral Trade Liberalization: Evidence for the United States," *American Economic Review*, 2006, *96* (3), 896–914.

_ **and Kamal Saggi**, "Tariff retaliation versus financial compensation in the enforcement of international trade agreements," *Journal of International Economics*, 2008, *76* (1), 48 – 60.

Maggi, Giovanni, "The Role of Multilateral Institutions in International Trade Cooperation," *American Economic Review*, 1999, *89* (1), 190–214.

Melitz, Marc J., "The Impact of Trade on Intra-Industry Reallocations and Aggregate Industry Productivity," *Econometrica*, 2003, *71* (6), 1695–1725.

Narayanan, Badri, Angel Aguiar, and Robert McDougall, eds, *Global Global Trade, Assistance, and Production: The GTAP 8 Data Base.* Center for Global Trade Analysis, Purdue University. http://www.gtap.agecon.purdue.edu (accessed October 2, 2014) 2012.

Ossa, Ralph, "Trade Wars and Trade Talks with Data," *American Economic Review*, 2014, *104* (12), 4104–46.

_ , "Why trade matters after all," *Journal of International Economics*, 2015, *97* (2), 266–277.

_ , "Quantitative Models of Commercial Policy," *National Bureau of Economic Research Working Paper Series*, 2016, *No. 22062.*

Park, Jee-Hyeong, "International trade agreements between countries of asymmetric size," *Journal of International Economics*, April 2000, *50* (2), 473–495.

Saggi, Kamal, Andrey Stoyanov, and Halis Murat Yildiz, "Do Free Trade Agreements Affect Tariffs of Nonmember Countries? A Theoretical and Empirical Investigation," *American Economic Journal: Applied Economics*, 2018, *10* (3), 128–70.

Su, Che-Lin and Kenneth L. Judd, "Constrained Optimization Approaches to Estimation of Structural Models," *Econometrica*, 2012, *80* (5), 2213–2230.

WTO, "Dispute Settlement Rules: Understanding on Rules and Procedures Governing the Settlement of Disputes," 1994. Marrakesh Agreement Establishing the World Trade Organization, Annex 2.

Figure 1: Trade-Weighted Tariffs in Japan

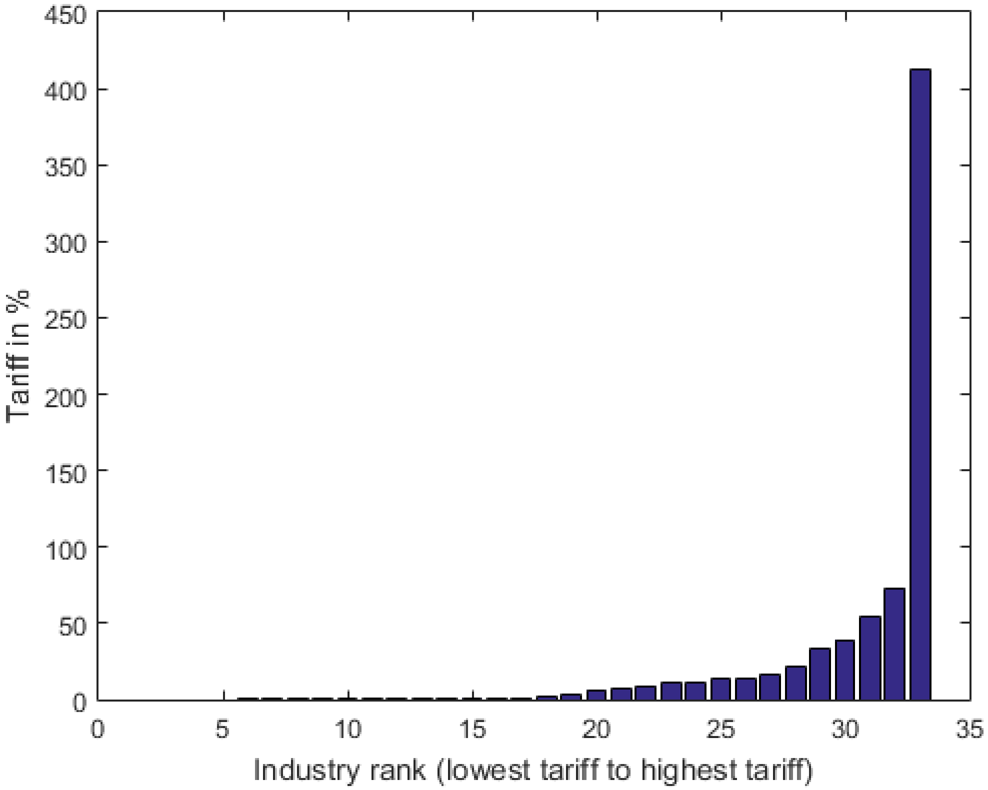
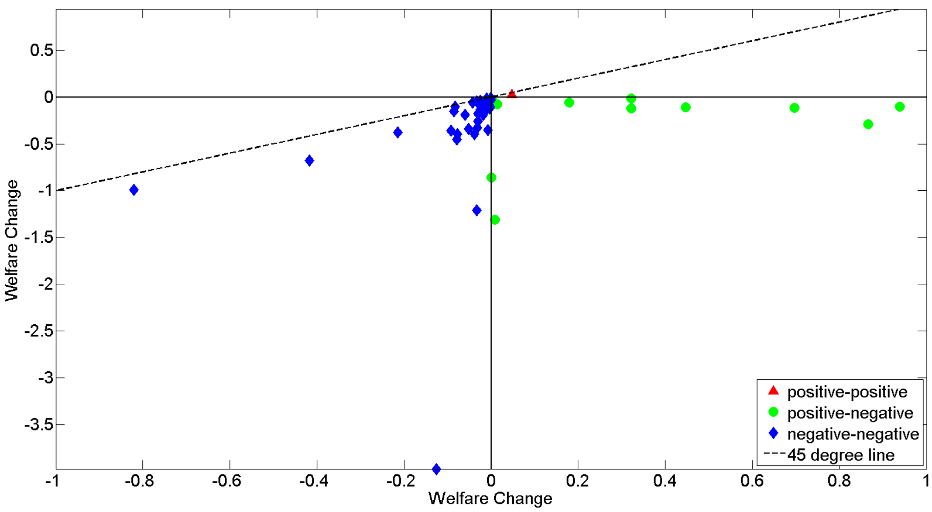
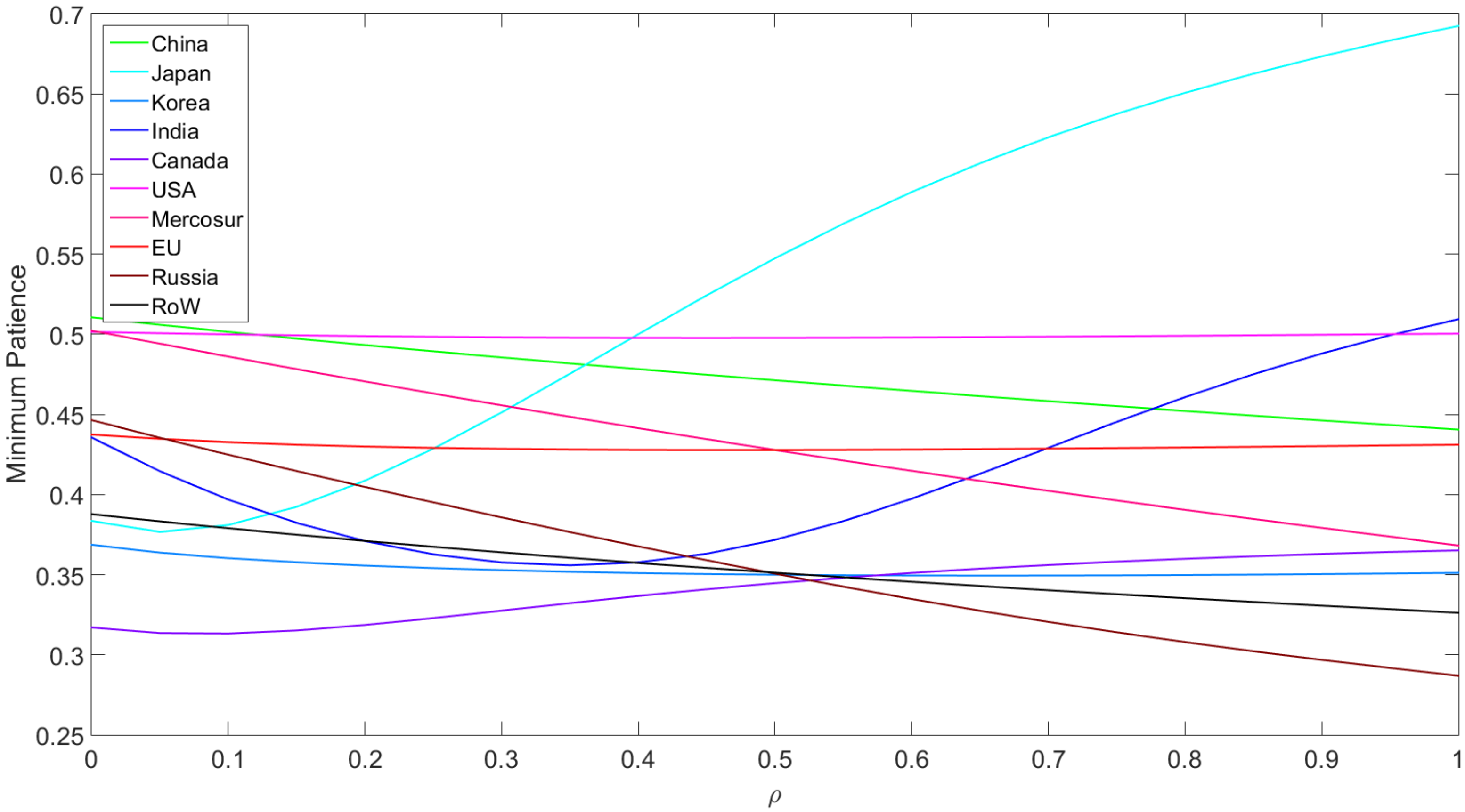


Figure 2: Percentage Welfare Changes in Bilateral Trade War



Notes: This scatter plot shows the percentage welfare changes in bilateral trade war. For each country pair, the country with the greater welfare change is always set to be the x-coordinate.

Figure 3: $\underline{\beta_j}$ with Counterfactual Negotiated Tariffs



Notes: This figure illustrates how the minimum patience of each country varies with world wide trade liberalization. Trade war is taken to be the static equilibrium in the punishment phase when calculating $\underline{\beta_j}$. ρ denotes the ratio of counterfactual negotiated tariffs to the factual tariffs, ranging from 0 (free trade) to 1 (factual tariffs).

Figure 4: Average Welfare Gain from Deviation with Trade Liberalization

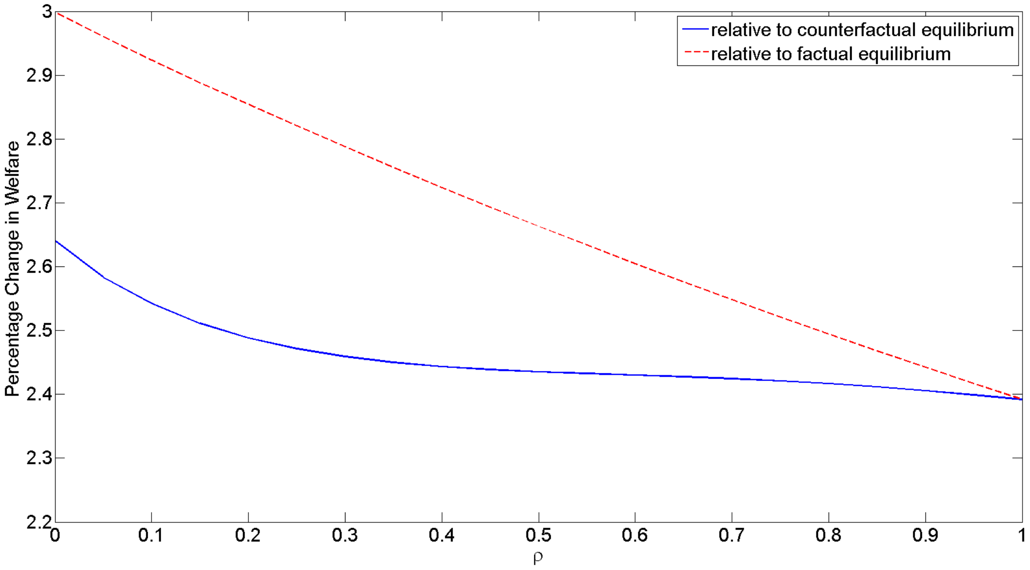


Figure 5: Average Welfare Loss in Punishment Phase with Trade Liberalization

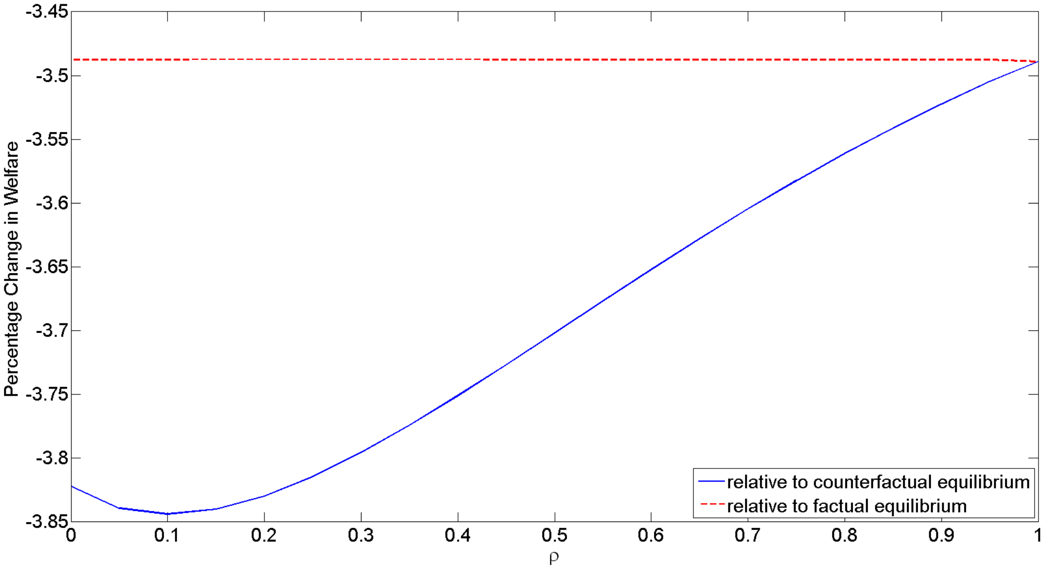
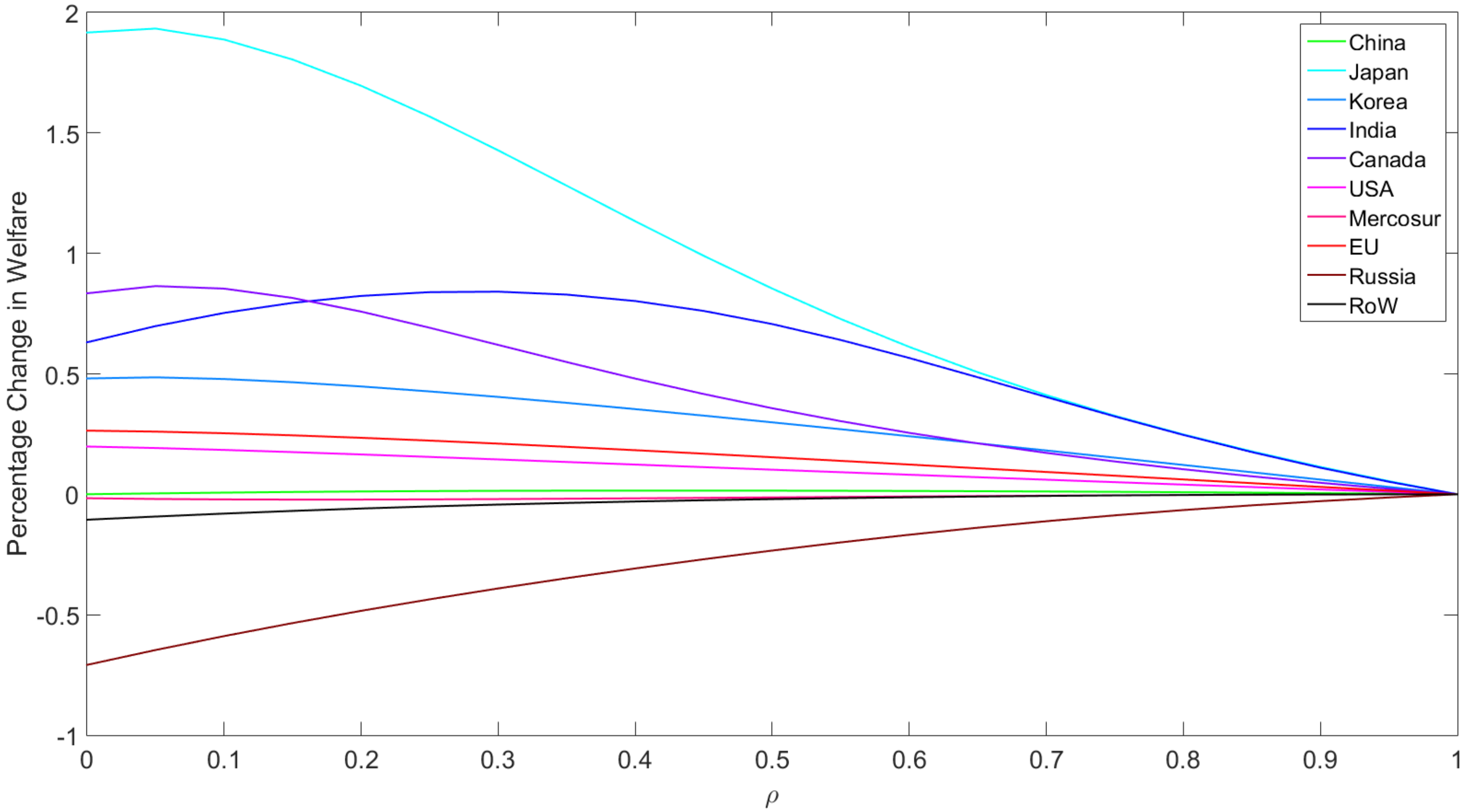


Figure 6: $\Psi_j(\rho)$ with Trade Liberalization



Notes: This figure illustrates how $\Psi_j(\rho)$ for each country varies with ρ.

Table 1: Summary by Sectors

	σ	Applied Tariff	Political Weight
Raw and processed rice	4.873	0.319	1.336
Wheat	12.373	0.106	1.443
Cereal grains nec	3.291	0.118	1.116
Vegetables, fruit, nuts	2.188	0.158	0.886
Oil seeds	2.889	0.18	1.32
Raw and processed sugar	2.521	0.228	1.054
Plant-based fibers	2.333	0.019	0.676
Crops nec	2.536	0.093	0.846
Cattle,sheep,goats,horses	2.582	0.031	0.864
Animal products nec	2.124	0.044	0.694
Raw and processed dairy	5.604	0.296	1.358
Wool, silk-worm cocoons	2.893	0.041	0.835
Forestry	2.332	0.028	0.741
Meat: cattle,sheep,goats,horse	4.391	0.114	1.147
Meat products nec	3.135	0.145	1.081
Vegetable oils and fats	4.981	0.112	1.153
Food products nec	2.777	0.141	0.993
Beverages and tobacco products	2.935	0.263	1.186
Textiles	2.903	0.1	1.014
Wearing apparel	5.311	0.123	1.247
Leather products	4.106	0.094	1.121
Wood products	2.288	0.066	0.82
Paper products, publishing	2.733	0.04	0.857
Chemical,rubber,plastic prods	2.375	0.05	0.816
Mineral products nec	2.47	0.065	0.871
Ferrous metals	3.014	0.044	0.957
Metals nec	4.384	0.042	1.041
Metal products	2.794	0.067	0.933
Motor vehicles and parts	3.126	0.086	0.986
Transport equipment nec	2.992	0.047	0.891
Electronic equipment	2.492	0.023	0.818
Machinery and equipment nec	2.374	0.047	0.831
Manufactures nec	3.515	0.065	1.068
Average	3.444	0.103	1

Notes: Entries under “σ” are estimates of elasticity of substitution. Entries under “Applied Tariff” are averages of applied tariffs. Entries under “Political Weight” are averages of estimated political weights which are scaled to have mean one.

Table 2: Welfare Changes in the Deviation Phase

	No Political Weights		With Political Weights	
	Own	Other	Own	Other
China	1.62	-0.89	1.2	-0.69
Japan	3.37	-0.37	2.55	-0.48
Korea	2.5	-0.17	2.9	-0.2
India	1.99	-0.06	0.59	-0.04
Canada	4.5	-0.11	3.44	-0.1
United States	2.38	-1.55	2.55	-1.53
Mercosur	1.1	-0.09	0.81	-0.06
European Union	1.89	-1.25	2.21	-1.38
Russia	1.88	-0.07	1.46	-0.05
Rest of World	2.68	-1.22	2.84	-1.21
Mean	2.39	-0.58	2.05	-0.57

Notes: The entries under "Own" are the percentage welfare change of the deviating country. Entries under "other" are the average of percentage welfare changes of other non-deviating countries.

Table 3: Welfare Changes in the Punishment Phase

	WTO Bound	Sanction	Trade War	Autarky
China	1.03	-0.29	-2.06	-11.6
Japan	3.03	0.59	-1.5	-12.97
Korea	2.09	-1.63	-4.62	-21.09
India	1.05	-0.21	-1.91	-10.98
Canada	3.27	-3.03	-7.82	-31.79
United States	1.5	-0.38	-2.38	-13.35
Mercosur	0.15	-0.23	-1.89	-9.93
European Union	1.09	-0.87	-2.5	-12.09
Russia	0.89	-1.56	-4.67	-20.31
Rest of World	1.59	-2.15	-5.55	-23.09
Mean	1.57	-0.98	-3.49	-16.72

Notes: Each entry represents the deviating country's percentage welfare change relative to the factual equilibrium in each period of the punishment phase.

Table 4: Welfare Changes in the Punishment Phase (With Political Weights)

	WTO Bound	Sanction	Trade War	Autarky
China	0.55	-1.07	-3.19	-13
Japan	2.24	0.65	-1.31	-11.23
Korea	2.56	-0.37	-3.83	-20.09
India	-0.41	-1.64	-3.61	-11.97
Canada	2.26	-4.05	-8.96	-32.61
United States	1.72	-0.1	-2.18	-12.93
Mercosur	-0.22	-1.07	-2.87	-10.7
European Union	1.42	-0.3	-2.1	-11.8
Russia	0.42	-2.33	-5.87	-21.13
Rest of World	1.72	-2.04	-5.8	-23.25
Mean	1.23	-1.23	-3.97	-16.87

Notes: Each entry represents the deviating country's percentage welfare change relative to the factual equilibrium in each period of the punishment phase.

Table 5: Median Tariffs in Sanction and Trade War

Median Tariffs	Factual		Sanction		Trade War	
	Own	Other	Own	Other	Own	Other
China	4.5	5.6	60.2	27.2	57.7	55.9
Japan	1.0	5.4	56.5	29.0	56.5	55.2
Korea	5.3	5.1	55.6	29.0	56.0	54.6
India	15.0	3.2	53.9	26.0	53.6	57.0
Canada	0.7	5.0	57.6	27.4	54.9	57.3
United States	0.9	5.1	62.2	28.6	59.4	55.9
Mercosur	8.2	4.3	56.0	25.9	54.7	56.7
European Union	2.1	6.1	57.5	27.6	57.9	56.5
Russia	8.8	2.2	59.8	27.0	54.5	55.6
Rest of World	6.3	3.2	61.1	28.8	59.6	57.1
Mean	5.3	4.5	58.0	27.7	56.5	56.2

Notes: The entries under "Own" are median tariffs imposed *against* other countries, whereas those under "Other" are median tariffs imposed *by* other countries. Note that under "Sanction," each row represents one static equilibrium with one specific deviating country, whereas under "Trade War," all entries are calculated from one static equilibrium.

Table 6: Minimum Patience

	WTO Bound	Sanction	Trade War	Autarky
China	2.76	0.85	0.44	0.12
Japan	9.78	1.21	0.69	0.21
Korea	6.04	0.6	0.35	0.11
India	2.12	0.91	0.51	0.15
Canada	3.65	0.6	0.37	0.12
United States	2.7	0.86	0.5	0.15
Mercosur	1.16	0.83	0.37	0.1
European Union	2.36	0.69	0.43	0.14
Russia	1.89	0.55	0.29	0.08
Rest of World	2.44	0.56	0.33	0.1
Mean	3.49	0.76	0.43	0.13

Notes: The entries are $\underline{\beta_j}$ calculated from equation (3) under four different scenarios in the punishment phase.

Table 7: Minimum Patience (With Political Weights)

	WTO Bound	Sanction	Trade War	Autarky
China	1.84	0.53	0.27	0.08
Japan	8.05	1.34	0.66	0.19
Korea	8.43	0.89	0.43	0.13
India	0.59	0.26	0.14	0.05
Canada	2.92	0.46	0.28	0.1
United States	3.07	0.96	0.54	0.16
Mercosur	0.79	0.43	0.22	0.07
European Union	2.8	0.88	0.51	0.16
Russia	1.4	0.39	0.2	0.06
Rest of World	2.55	0.58	0.33	0.11
Mean	3.24	0.67	0.36	0.11

Notes: The entries are $\underline{\beta}_j$ calculated from equation (3) under four different scenarios in the punishment phase.

Table 8: Percentage Welfare Changes with Unilateral Country-Specific Tariffs

	Self			Target			Others		
	Mean	Max	Min	Mean	Max	Min	Mean	Max	Min
China	0.09	0.34	0.01	-0.5	-0.14	-1.46	0.01	0.29	-0.03
Japan	0.44	1.35	0.01	-0.17	-0.04	-0.36	0	0.03	-0.01
Korea	0.13	0.29	0.01	-0.08	0.01	-0.23	0	0.03	-0.01
India	0.30	0.75	0.02	-0.06	-0.02	-0.12	0	0.01	-0.01
Canada	0.32	1.60	0	-0.06	-0.01	-0.38	0	0.03	0
United States	0.13	0.37	0	-0.91	-0.14	-5.29	0.02	0.3	-0.05
Mercosur	0.06	0.16	0.01	-0.05	-0.01	-0.09	0	0.01	0
European Union	0.11	0.47	0.02	-0.64	-0.30	-1.66	0.02	0.17	-0.1
Russia	0.11	0.53	0	-0.03	0	-0.10	0	0.01	-0.03
Rest of World	0.14	0.67	0.02	-0.71	-0.31	-1.49	0.04	0.58	-0.01
Mean	0.18	0.65	0.01	-0.32	-0.1	-1.12	0.01	0.15	-0.02

Notes: The entries under “Self” are summary statistics of welfare changes of countries imposing optimal tariffs against one target country only. The entries under “Target” are summary statistics of welfare changes of the target countries. The entries under “Others” are summary statistics of welfare changes of the countries that are neither the deviating country nor the target country. Note that for each country, the “Self” and “Target” columns each contain a sample of size 9, whereas the “Other” column contains a sample of size 72.

Table 9: Winners and Losers of Bilateral Trade War

(Winner, Loser)	Average Tariffs		Import Share		Export Share	
	Winner	Loser	Winner	Loser	Winner	Loser
(China, Korea)	0.06	0.18	0.12	0.21	0.07	0.36
(China, RoW)	0.06	0.08	0.39	0.13	0.24	0.22
(Japan, China)	0.26	0.05	0.25	0.16	0.27	0.15
(Japan, USA)	0.22	0.04	0.18	0.08	0.15	0.09
(Japan, RoW)	0.23	0.08	0.32	0.09	0.27	0.1
(India, Canada)	0.26	0.04	0.01	0.01	0.01	0.01
(India, Mercosur)	0.26	0.09	0.01	0.01	0.02	0.01
(India, Russia)	0.17	0.09	0.02	0.01	0.01	0.02
(Canada, EU)	0.1	0.06	0.1	0.02	0.1	0.02
(Mercosur, Korea)	0.09	0.1	0.03	0.01	0.02	0.01
(Mercosur, Russia)	0.05	0.11	0.02	0.02	0.02	0.02
Mean	0.16	0.08	0.13	0.07	0.11	0.09

Notes: Entries under “Average Tariffs” are the winner or loser country’s average tariff level against each other. Entries under “Import share” are the total imports of the winner or loser from the other country as the share of total imports. Entries under “Export share” are calculated similarly.

Table 10: Multilateral versus Bilateral Enforcement Mechanism

	Welfare			Sustainability	
	ΔW	deviating	deviated	$\underline{\beta_j}^{BI}>\underline{\beta_j}^{M}$	$\underline{\beta_j}^{BI}<\underline{\beta_j}^{M}$
China	0.03	2	1	2	5
Japan	0.80	3	0	3	3
Korea	-0.94	1	3	4	4
India	0.80	4	1	3	2
Canada	0.46	1	1	2	6
United States	-0.01	0	1	3	6
Mercosur	0.04	2	1	3	4
European Union	0.10	0	1	4	5
Russia	-0.08	0	2	3	6
Rest of World	-1.08	0	2	5	4

Notes: The three columns under "Welfare" are the resulting percentage welfare changes in the new equilibrium when both sides of the unsustainable country pair impose bilateral Nash tariffs, and the number of cases in which the country deviates or is deviated against. The remaining two columns describe changes in the propensity to deviate for the remaining country pairs. Entries under "$\underline{\beta_j}^{BI} > \underline{\beta_j}^{M}$" and "$\underline{\beta_j}^{BI} < \underline{\beta_j}^{M}$"are the respective number of cases in which the calculated minimum patience is greater/less than that if the corresponding multilateral enforcement were used in the punishment phase.

Table 11: Complementarity Effect

	ΔW			Median Tariffs		
	FTA	non-FTA	Factual	FTA	non-FTA	Factual
China	1.22	1.62	1.62	44.9	58.1	58.0
Japan	2.96	3.37	3.37	43.2	57.6	57.3
Korea	1.86	2.50	2.50	41.0	56.1	56.1
India	1.64	1.98	1.99	44.1	54.1	53.5
Canada	3.69	4.49	4.50	44.0	55.5	55.5
United States	1.86	2.38	2.38	45.6	60.3	60.3
Mercosur	0.82	1.10	1.10	43.7	54.7	54.6
European Union	1.49	1.89	1.89	45.3	59.5	59.5
Russia	1.36	1.87	1.88	42.9	53.0	52.8
Rest of World	1.96	2.68	2.68	45.9	60.4	60.4
Mean	1.89	2.39	2.39	44.1	56.9	56.8

Notes: Entries under "ΔW" are the average percentage changes in welfare in the deviation phase, measured relative to the factual equilibrium. Entries under "Median Tariffs" are the average of median optimal tariffs imposed in the deviation phase. Columns labeled "FTA", "non-FTA" and "Factual" refer to the stage equilibrium at the beginning of the repeated game: counterfactual equilibria in which the country is a FTA member, those in which the country is not a member, and the factual equilibrium respectively. Entries under "FTA" are average values over nine equilibria while those under "non-FTA" are averages over 36 equilibria. The last row reports averages over countries.

Table 12: Punishment Effect

	ΔW			Median Tariffs		
	FTA	non-FTA	Factual	FTA	non-FTA	Factual
China	-1.82	-1.93	-2.06	45.6	57.5	57.7
Japan	-1.00	-1.35	-1.50	43.5	56.0	56.5
Korea	-3.84	-4.40	-4.62	41.4	55.6	56.0
India	-1.44	-1.80	-1.91	45.0	53.7	53.6
Canada	-6.58	-7.42	-7.82	45.0	54.9	54.9
United States	-2.11	-2.24	-2.38	46.3	59.1	59.4
Mercosur	-1.58	-1.77	-1.89	44.4	54.9	54.7
European Union	-2.20	-2.38	-2.50	45.6	57.6	57.9
Russia	-4.03	-4.39	-4.67	45.4	55.2	54.5
Rest of World	-5.07	-5.32	-5.55	47.0	59.4	59.6
Mean	-2.97	-3.30	-3.49	44.9	56.4	56.5

Notes: Entries under "ΔW" are the one-period average percentage changes in welfare in the punishment phase, measured relative to the factual equilibrium. Entries under "Median Tariffs" are the average of the median optimal tariffs imposed in the punishment phase. Columns labeled "FTA," "non-FTA," and "Factual" refer to the stage equilibrium at the beginning of the repeated game: counterfactual equilibria in which the country is an FTA member, those in which the country is not a member, and the factual equilibrium, respectively. Entries under "FTA" are average values over nine equilibria, whereas those under "non-FTA" are averages over 36 equilibria. The last row reports averages over countries.

Table 13: Minimum Patience with FTA

	FTA	non-FTA	Factual
China	0.39	0.46	0.44
Japan	0.64	0.72	0.69
Korea	0.32	0.36	0.35
India	0.51	0.53	0.51
Canada	0.38	0.38	0.37
United States	0.46	0.52	0.5
Mercosur	0.34	0.39	0.37
European Union	0.39	0.44	0.43
Russia	0.26	0.3	0.29
Rest of World	0.27	0.34	0.33
Mean	0.40	0.44	0.43

Notes: Columns labeled "FTA," "non-FTA," and "Factual" refer to the stage equilibrium at the beginning of the repeated game: counterfactual equilibria in which the country is an FTA member, those in which the country is not a member, and the factual equilibrium, respectively. Entries under "FTA" are the country's average $\underline{\beta}_j$ over nine equilibria, whereas those under "non-FTA" are averages over 36 equilibria. The last row reports averages over countries.