The Principle of Reciprocity in the 21st Century

David R. DeRemer*
International School of Economics, Kazakh-British Technical University

May 2019

Abstract

In modern trade environments with offshoring and politically powerful firms, governments have reason to use trade agreements to expand export volumes beyond what would be expected from standard trade agreement models. In such a setting, we characterize the outcome of rules-based trade negotiations according to WTO principles. The outcome yields new predictions that are consistent with empirical evidence concerning cooperative tariffs under the WTO and governments' attempts to deviate from the cooperative equilibrium of the WTO. Exporters with greater political power and larger supply elasticities compel greater reductions in cooperative import tariffs through trade negotiations — reductions so large that losses to domestic firms from import competition outweigh gains to consumers. Unlike the standard model's cooperative equilibrium in which a government can gain only by improving its terms of trade, a government here will seek to improve outcomes for the losing import-competing firms by imposing disguised protection and discouraging foreign export subsidies.

JEL Codes: F12, F13, F15, L24

Keywords: trade agreements, principle of reciprocity, WTO.

*E-mail: d_deremer@ise.ac. Web: www.nber.org/~dderemer. Kyle Bagwell, Emily Blanchard, Paola Conconi, Meredith Crowley, Xenia Matschke, Ralph Ossa, Robert Staiger, seminar participants at Trier University and conference participants at the Stanford Institute for Theoretical Economics, DEGIT, European Trade Study Group, FIW, InsTED and MWIEG provided valuable comments. Essential financial support was provided by the MTA Lendület program, the Columbia NSF-IGERT International Globalization and Development Program, and the project "Dispute Settlement in Trade: Training in Law and Economics" (DISSETTLE), a Marie Curie Initial Training Networks (ITN) Funded under the EU's Seventh Framework Programme, Grant Agreement No. FP7-PEOPLE-2010-ITN_264633.

1 Introduction

This paper is about trade negotiation outcomes in modern international economies. The focus of trade negotiations has shifted in recent years from import tariffs in final goods to tariff and non-tariff barriers facing intermediate goods in global supply chains (see e.g. Baldwin, 2013, 2016). An alternative perspective is that the role of modern trade agreements is to facilitate rent-seeking by politically powerful exporters (Rodrik, 2018). These shifts raise a number of questions for the political economy of trade agreements. What political-economic problems are trade agreements solving? How well do prevailing principles of trade negotiations and agreements address these problems? Can we predict and confirm how trade negotiation outcomes differ in this new era, and how countries will attempt to renege on trade agreements? To address these questions, this paper applies the long-accepted principle of reciprocity in trade negotiations to develop new predictions for trade agreement outcomes in modern settings. We then explore what insights these predictions can yield into observed trade cooperation, ranging from conventional import tariff cuts to the present struggles of the World Trade Organization (WTO).

The main contribution of this paper is a new answer to the question, what incentives do governments have to manipulate local prices in their markets after signing trade agreements? In the simplest trade agreement models — including the standard model of nations maximizing national income (Johnson 1953-54) and generalizations to political economy under perfect competition (Bagwell and Staiger, 1999) — the answer is none. In these frameworks, countries will engage in reciprocal liberalization until governments value no change in local prices, so the only first-order gain a country can achieve through policy changes is by improving terms of trade at the other nation's expense: the resulting outcome is what Bagwell and Staiger call the *political optimum*. In the current paper, governments will engage in deeper liberalization to further promote exporters at the expense of import-competing firms. Governments will engage in such behavior if there is imperfect competition and an absence of negotiations of export policies, as is the case in the WTO. After a trade agreement, governments will then be tempted to engage in protection that benefits import-competing firms, even if such policies have no effect on tariff revenue or their terms of trade. As we will detail, such a trade agreement outcome matches better with the recent experience of the WTO. Neither the recent descent into disguised protectionism following the global financial crisis of 2008 nor the blatant protectionism started under the Trump administration in 2017 can be explained well by the terms-of-trade motive (Crowley, 2019), but such deviations are consistent with the cooperative equilibrium that we define.[1]

In considering trade negotiation outcomes, this paper builds on the general trade agreement model of two governments with static preferences over equilibrium price outcomes, which in turn depend on policy choices (Bagwell and Staiger, 1999). Here prices include both the local prices prevailing in each market and the world prices i.e. the terms of trade, that prevail in between borders. The emphasis of this paper is that international externalities due to changes in local prices will matter at trade agreements outcomes, and it is non-trivial to devise a trade agreement model with this result. In practice, changes in local prices will affect foreigners at trade agreement outcomes because they affect foreigners' producer surplus or value-added. This abstraction of international *local price externalities* covers a variety of potential problems that nations can address through trade negotiations. For example, an individual government's policy tradeoffs may account for benefits to domestic final good producers but ignore surplus flowing upstream to intermediate producers abroad. Another example would be a government imposing import tariffs while ignoring their effect on politically-organized exporters abroad. Negotiations involving local price externalities thus permit the wider role for politics in trade agreements that Rodrik (2018) observes. Satisfactory modelling of such negotiations is non-trivial, because in many standard models, trade agreements need only address *terms-of-trade externalities* : the ability for countries to exploit their market power and manipulate world prices to improve their national income at their trading partners' expense. Though local price externalities are now better understood and becoming more important, the appropriate design of trade agreements and trade institutions in their presence has received limited formal attention.

Why have local price externalities become more important for trade negotiations over time? The relevance of local price externalities is a consequence of nations lacking policies to address these concerns unilaterally, and recent literature offers reasons why unilateral policy options are increasingly limited. First, when firms offshore and intermediate prices are determined through bilateral bargaining, the nation hosting intermediate exporters can lack the ability to influence the final good price in the nation downstream (Antràs and Staiger, 2012a, 2012b; Blanchard, Bown, and Johnson, 2016). Second, the World Trade Organization (WTO) since 1995 has firmly institutionalized limitations on export subsidies and domestic subsidies. Absent the ability to impose such subsidies, additional trade agreement problems can arise, such as longer-run firm delocation (Ossa, 2011; Bagwell and Staiger, 2015) and

[1] Specifically, the eBook of Crowley (2019) surveys research describing how tariffs imposed by the Trump administration have been fully passed on to importers, hence the tariffs do not provide any terms-of-trade benefits for the importing countries imposing the tariffs. Disguised protectionism (from Global Trade Alert) described in the volume often involves red-tape barriers for import-competing sectors that cannot easily be explained by passing policy costs off to foreigners, as would be necessary for the terms-of-trade explanation.

shorter-run profit-shifting (Mrázová, 2011; Ossa, 2012; Bagwell and Staiger, 2012b). Such problems can arise even in a perfect competition model with short-run profits and sectoral differences in political influence (Bagwell and Staiger, 2016).

Given the importance of local price externalities, this paper develops a unified approach for predicting trade agreement outcomes in modern settings where these externalities matter. A new prediction is valuable, because the leading approach for deriving cooperative outcomes fails to yield efficient policies for several models surveyed by Bagwell and Staiger (2016). The approach that fails is the derivation of politically optimal policies—those nations choose when they act as if they do not value terms-of-trade manipulation. The current paper instead builds on the Bagwell and Staiger theory that what guides nations to cooperative outcomes is the norm in rules-based trading systems of reciprocal policy changes—those which increase trade volume equally when valued at world prices. In selecting among possible trade agreement outcomes on the Pareto efficiency frontier, *stability with respect to reciprocal changes* is an appealing criterion because (1) it reflects an end point of gradual reciprocal negotiations such that neither nation can gain further from reciprocal changes, and (2) it relates to norms of renegotiation and retaliation codified in the 1947 General Agreement on Tariffs and Trade (GATT), in that no nation gains from a unilateral policy change when trading partners can respond reciprocally. Bagwell and Staiger find that this stability property holds for a variety of settings in which the politically optimal policies are efficient. The current paper builds on this finding by evaluating more generally how stable policies under reciprocity can achieve efficiency. Such stable policies can serve as suitable predictions for cooperative outcomes even when efficient trade agreements must address local price externalities.

With new predictions in hand, we can then evaluate whether the stable agreements that address local price externalities lend new insight into observed trade cooperation. We highlight here a few important observations that are not easily rationalized when trade agreements address only terms-of-trade externalities. One is that politically-organized exporters can induce lower levels of trading partners' import tariffs, as shown for the 1986-1994 Uruguay Round negotiations by Ludema and Mayda (2013). Such a result also relates to a practitioner's conventional wisdom from Regan (2015) that negotiations can cut into import tariffs that have purely distributional motives and no terms-of-trade benefits. Another observation is that at cooperative outcomes, governments make additional choices that increase local prices in import-competing sectors yet do not necessarily yield terms-of-trade gains, such as unilateral non-tariff barriers and multilateral subsidy restrictions. Such policies are less straightforward to explain when agreements address only terms-of-trade externalities, because (as in Bagwell and Staiger, 1999) each government benefits from local price decreases in the import-competing sector along the liberalization path, up until the stable and efficient outcome where there are no first-order effects from local price changes. An important question then is whether local price externalities lead to the new possibility that at the stable negotiation outcome, each government would benefit from higher local prices in import-competing sectors. In addition to interpreting recent trade policy history, we also apply the theory to develop new predictions for government cooperation in offshoring settings.

To begin the formal inquiry, Section 2 develops the framework and technical tools for using stability under reciprocity to select among Pareto efficient trade policies. We define a general model of trade agreements addressing international externalities, and we define noncooperative policies and the Pareto efficiency frontier. We propose stability under reciprocity rules as a general approach to select among outcomes on the efficiency frontier. We define general properties of reciprocity rules, and we show how stability under reciprocity can imply efficiency. To aid in deriving stable policies, we show that they can be represented as an equilibrium when nations unilaterally optimize subject to a reciprocity constraint. The constrained optimization equilibrium approach generalizes a theory of deriving efficient policies subject to market access constraints (Bagwell and Staiger, 2001a). The theory here provides the foundation that justifies stability under reciprocity as an appropriate concept for selecting an efficient trade agreement outcome even when local price externalities matter.

Section 3 adopts a specific rule, the standard reciprocity rule, such that reciprocal policy changes increase the value of trade equally at prevailing world prices. We add structure that implies policies affect government objectives through local prices, world prices, and potentially foreign local prices. To demonstrate the application of the standard reciprocity rule, we first apply it to the simpler case where trade agreements can achieve efficiency by eliminating terms-of-trade manipulation. We do this both to justify the focus on the standard reciprocity rule, as well as to clarify the circumstances that imply trade agreements must address local price externalities. We find that if nations' policy spaces exhibit specific forms of completeness, trade agreements need only address terms-of-trade externalities. Completeness holds either if nations can achieve a first-best allocation of production, or if each nation has policies to affect every price that matters for trade. We affirm under such conditions, stable and efficient policies exist. The failure of these conditions is then necessary for local price externalities to matter for trade agreements.

The focus of Section 4 is on generalizing stability under reciprocity to develop predictions for trade agreement outcomes when local price externalities do matter. A limitation of the approach is that stable outcomes under reciprocity do not always exist. Existence fails if across nations there are redundant policies for targeting some local prices but not others. We find though that the existence of stable outcomes is facilitated by allowing for reciprocal negotiations over a restricted policy space, and this restriction comes at no cost to efficiency. One possible approach for traditional trade in final goods is to allow reciprocal negotiations over only import tariffs while also preventing domestic policy changes that could substitute for tariffs in limiting market access. Historical GATT/WTO norms and rules match this approach in permitting reciprocal negotiating over import tariffs and not over export policies. We find that the efficient and stable outcome under reciprocity over the limited policy space unusually allows for the possibility that nations benefit from local price increases in import-competing sectors. This possibility helps to explain a number of empirical facts, which we can illustrate by proceeding to consider more specific trade environments.

The first of the examples that we consider in Section 5 is the simple partial equilibrium model of Bagwell and Staiger (2001b) with a freely-traded numeraire good, two other costly-traded goods, and politically-organized importers and exporters. Following Bagwell and Staiger (2016), when government policies are limited to import taxes, the model satisfies our criteria for the importance of local price externalities for trade agreements. There are no domestic policies that correct the political distortion between the import sector and export sector. Nations care about the export price, and they have no export policy to address this concern. We confirm in this setting that the stable outcome with respect to reciprocity is efficient. Trade volumes at the stable outcome are still nonetheless the same as the stable outcome if nations had both import and export policies.

At this stable efficient point for the limited-instrument, partial equilibrium model with politically-organized exporters, we find that nations benefit from a higher price in the import-competing sector. In order to gain more profits for the exporters, governments offer additional liberalization in the import-competing sectors—so much so that the losses to the import-competing industry are larger than the gains to consumers from the marginal liberalization. Governments would then benefit from introducing either non-tariff measures or temporary trade barriers that increases prices in import-competing sectors, even when there is no motive for terms-of-trade manipulation. Rules to limit export policies take the form of bans on export subsidies if the loss from further local price decreases dominates the terms-of-trade gains from such subsidies.

Our second example builds on the offshoring model of Antràs and Staiger (2012b), in which the Home nation is a final good producer and the Foreign nation values the rents accruing to its intermediate export industry. Home has a policy that affects the price for the final good and demand for the intermediate, but sets the final good price too low for global efficiency because it does not internalize rents accruing to the intermediate producers. The Foreign government can impose a noncooperative tax on the intermediate to increase rents and improve its terms of trade. Like in Antràs and Staiger, Foreign lacks an instrument to affect the final good price. If Home also has an import tax on the intermediate, then there is no stable and efficient result of negotiations. But negotiations can still achieve efficiency when Home fully liberalizes imported intermediates and then there is reciprocity between Home's policy affecting the final good and Foreign's policy affecting intermediate exports. The stable outcome roughly matches the kind of liberalization discussed in Baldwin (2013, 2016) where there is full liberalization in barriers for the imported intermediate and then reciprocity between investment from final good producers in developed countries and reform from developing countries.

The last example we consider is a symmetric version of the offshoring model from Antràs and Staiger (2012b) with free trade in intermediates and reciprocal negotiations over final good tariffs. The policy space follows the global supply chain theory and empirics of Blanchard, Bown, and Johnson (2016). Our results imply that there are relevant local price externalities from the final good tariffs, and our model predicts that bilateral agreements should address these externalities. One additional prediction is that well-integrated nations subject to external import competition could cooperate in raising external barriers to trade. This is bilateral opportunism to promote internal trade in intermediates and more shared rents from protecting non-traded final goods, rather the standard bilateral opportunism involving changes in internal trade barriers for final goods (Bagwell and Staiger, 2005a). This prediction relates to concerns about trade diversion from cooperative non-tariff barriers in preferential trade agreements (e.g. Limão, 2016) and the common talking point that mega-regionals help nations that become "rule-makers" in the global trading system. This concludes our demonstration of the new theory's scope of application.

The closest related studies are Ossa (2011), Mrázová (2011), Bagwell and Staiger (2016), and Cole, Lake, and Zissimos (2018), because they each consider reciprocity in trade agreement models that feature local price externalities. Ossa and Mrázová each consider import policies for a specific imperfectly-competitive market structure and propose alternative reciprocity rules that preserve sectoral trade balance in their models.[2] Bagwell and Staiger (2016) consider the current paper's definition of reciprocity in the full-instrument version of Antràs and Staiger (2012b) and the limited-instrument version of Bagwell and Staiger (2001b). They observe in both settings that the political optimum is neither efficient nor stable, and they do not propose any alternative prediction. Cole, Lake, and Zissimos develop a model of trade agreement formation that emphasizes the political fight between interest groups over trade agreement ratification. Their model features a local price externality (namely, a political externality), because the foreign tariff, via the foreign local price, affects the probability of

[2] We discuss alternative rules further in Section 3 and the conclusion. Another difference from Ossa is that the cooperative equilibrium is a corner solution rather than an interior solution (see footnote 8).

trade agreement ratification. The efficient trade agreement could involve either more or less reciprocal liberalization than Bagwell and Staiger's political optimum.

Another related paper by Ludema and Mayda (2013) also derives a trade agreement outcome in which exporter political organization matters for the level of liberalization. They obtain this result in a perfect-competition, partial-equilibrium model with political economy, and they use Nash bargaining and a principal supplier rule to determine the trade agreement outcome, and they empirically identify how terms-of-trade externalities impact the outcome.[3]

In most related literature, terms-of-trade manipulation is the sole international externality that efficient trade agreements need to address. Particular models in this literature offer alternative explanations for each of the facts interpreted within the current paper's framework. Staiger and Sykes (2011) and Staiger (forthcoming) develop theory in which nations impose non-tariff measures unilaterally in order to shift costs of pursuing regulatory goals onto trading partners. To motivate export subsidy bans at stable and efficient policies, Bagwell and Staiger (2012a) and Bagwell and Lee (2015) each consider a particular market structure with long-run firm delocation such that nations can unusually improve their terms of trade by imposing export subsidies (i.e. they obtain Metzler paradoxes). Temporary trade barriers can be explained by exogenous shocks in political motives (Bagwell and Staiger, 2005b) or import shocks when the terms-of-trade gains from temporary barriers outweigh the price distortions (Bagwell and Staiger, 1990). Among the models where terms-of-trade externalities are the sole focus of trade agreements, there is no explanation for agreements reducing import tariffs imposed for purely redistributional motives, but such reductions do result when trade agreements address both terms-of-trade manipulation and the commitment problem of governments facing a domestic lobby (Maggi and Rodriguez-Clare, 2007).

To summarize what follows, Section 2 defines stability under reciprocity in a general framework, Section 3 introduces the standard reciprocity rule and its typical applications, Section 4 considers stability under reciprocity when local price externalities matter, Section 5 applies the new theory, and Section 6 concludes.

2 Modeling Trade Agreements and Reciprocity

This section develops the framework necessary to justify stability under reciprocity as a general concept for predicting trade agreement outcomes. First, we develop a general

[3] In Section 5.2, a new finding is that exporter political economy creates a local price externality that affects the tariff level at the negotiation outcome, but not at the Nash equilibrium tariffs. So Ludema and Mayda (2013) do not interpret this effect of exporter political organization as trade agreements addressing a local price externality.

two-country model of trade agreements that address international externalities. We then define reciprocity rules, and stability under reciprocity rules. We show that efficient and stable policies can be derived as an equilibrium that jointly solves two problems of unilateral constrained optimization subject to a reciprocity constraint.

2.1 A General Model of Agreements and Externalities

Consider two countries, Home and Foreign, who choose policy vectors Λ for Home and Λ^* for Foreign. These vectors include trade policies and possibly also domestic policies. Government objectives can be expressed as functions of policies $W(\Lambda, \Lambda^*)$ and $W^*(\Lambda, \Lambda^*)$, All objectives are continuous, differentiable, and globally concave within the policy space.

We define the *noncooperative equilibrium* according to *Nash equilibrium policies* Λ^N and Λ^{*N} that jointly satisfy the $|\Lambda| + |\Lambda^*|$ first-order conditions from unilateral optimization

$$\frac{dW}{d\Lambda} = 0 \text{ and } \frac{dW^*}{d\Lambda^*} = 0. \tag{1}$$

Here $\frac{dW}{d\Lambda}$ is a $1 \times |\Lambda|$ Jacobian matrix of derivatives, and we maintain similar notation for multivariate derivatives throughout the paper. We assume there are international externalities in all policy choices at the Nash equilibrium. Without loss of generality, we define the direction of policies so that at the Nash equilibrium,

$$\frac{dW}{d\Lambda^*} < 0 \text{ and } \frac{dW^*}{d\Lambda} < 0. \tag{2}$$

The vector inequality here is satisfied for all elements, and we maintain this convention for the inequality notation throughout the paper.

We define the set of *cooperative equilibria* on the *global efficiency frontier* for policies maximizing the objective $W + \lambda W^*$ for some positive multiplier λ. Throughout this paper we use "efficiency" to refer to Pareto efficient policies with respect to national preferences, and we use "unilaterally optimal" to refer to any maximization of a single government's objective.

Given a policy space (Λ, Λ^*), we denote the Pareto efficiency frontier within the policy space as $EF(\Lambda, \Lambda^*)$. If W and W^* are strictly concave, then every point on EF can be written as $(\hat{\Lambda}(\lambda), \hat{\Lambda}^*(\lambda))$ for some positive multiplier λ such that the objective $W + \lambda W^*$ is maximized and the following first-order conditions are satisfied:

$$\frac{dW}{d\Lambda} + \lambda \frac{dW^*}{d\Lambda} = 0 \text{ and } \frac{dW}{d\Lambda^*} + \lambda \frac{dW^*}{d\Lambda^*} = 0. \tag{3}$$

The efficiency conditions are then equivalent to establishing

$$\frac{dW}{d\Lambda_i} / \frac{dW^*}{d\Lambda_i} = \frac{dW}{d\Lambda_j} / \frac{dW^*}{d\Lambda_j} = -\lambda < 0 \tag{4}$$

for all policy pairs $(\Lambda_i, \Lambda_j) \in \{\Lambda \cup \Lambda^*\}^2$ when evaluated at $(\hat{\Lambda}(\lambda), \hat{\Lambda}^*(\lambda))$. If instead W and W^* are quasilinear (so not strictly concave), then there exist policy changes that are equivalent to transfers. Policies on EF then maximize the objective $W + W^*$, and EF is linear with a slope of -1 on the (W, W^*) plane.[4] By (1), (2) and (3), we can conclude that the Nash equilibrium $(\Lambda^N, \Lambda^{*N}) \notin EF$.

A fundamental question for the theory of trade agreements is which point on EF will be chosen following trade negotiations. In a symmetric case, we would anticipate the efficient point corresponding to the global objective with $\lambda = 1$ and $W = W^*$ and there is no problem left for theory to solve. For asymmetric cases we need additional assumptions to guide the selection of a point on EF.

In selecting a point on the efficiency frontier, stability under reciprocity has desirable properties. Following Bagwell and Staiger (2002, Ch. 2), one possible alternative would be a "power-based" Nash bargaining solution, but they argue that a "rules-based" system following the principle of reciprocity fits better with the history of the GATT/WTO. Bagwell and Staiger (1999) further specify a model in which we can also interpret stable and efficient policies as being robust to renegotiations under GATT/WTO rules. We proceed to develop general theory on how stability under reciprocity implies a point on the efficiency frontier.

2.2 General Reciprocity Rules

We define a *reciprocity rule* as a function $R(\Lambda, \Lambda^*; \Lambda^0, \Lambda^{0*}) = 0$ that specifies a range of permissible policies based on prevailing policies $(\Lambda^0, \Lambda^{0*})$. We further assume that R is monotonic and that the rule is satisfied at the status quo, so $R(\Lambda^0, \Lambda^{0*}; \Lambda^0, \Lambda^{0*}) = 0$. For any two policy pairs $(\Lambda_i, \Lambda_j) \in \{\Lambda \cup \Lambda^*\}^2$, differential policy changes $\frac{d\Lambda_j}{d\Lambda_i}$ satisfy the reciprocity rule at $(\Lambda^0, \Lambda^{0*})$ if and only if

$$-\frac{d\Lambda_j}{d\Lambda_i} = \frac{\frac{dR}{d\Lambda_i}}{\frac{dR}{d\Lambda_j}}. \tag{5}$$

We further assume that for Home policies, $R < 0$ is not permitted, and for Foreign policies, $R > 0$ are not permitted. For the rule then to act as a constraint on behavior causing negative externalities given (2), we must assume that R satisfies

[4] All models in Bagwell and Staiger (2012b, 2015) and Antràs and Staiger (2012a,b) have quasilinear preferences and policy changes that are equivalent to income transfers (local-price-preserving changes in import and export policies that affect the same price), so the EF is linear with slope -1.

$$\frac{dR}{d\Lambda} < 0 \text{ and } \frac{dR}{d\Lambda^*} > 0. \tag{6}$$

For any two policy pairs $(\Lambda_i, \Lambda_j^*) \in \Lambda \times \Lambda^*$, our assumption (6) implies that $\frac{d\Lambda_j^*}{d\Lambda_i} > 0$.

Before discussing stability, we verify that reciprocal policy changes can imply Pareto improvement starting from inefficient Nash equilibrium policies. Define that *nations gain from reciprocal policy reduction* at policies $(\tilde{\Lambda}, \tilde{\Lambda}^*)$ under reciprocity rule R if for all $(\Lambda_i, \Lambda_j^*) \in \Lambda \times \Lambda^*$ evaluated at $(\tilde{\Lambda}, \tilde{\Lambda}^*)$,

$$\begin{aligned} \frac{dW}{d\Lambda_i} + \frac{dW}{d\Lambda_j^*}\frac{d\Lambda_j^*}{d\Lambda_i} &< 0, \text{ and} \\ \frac{dW^*}{d\Lambda_i} + \frac{dW^*}{d\Lambda_j^*}\frac{d\Lambda_j^*}{d\Lambda_i} &< 0, \end{aligned} \tag{7}$$

where the sign is negative because we are considering small reductions in Λ_i and Λ_j. We can easily verify that nations gain from reciprocal policy reductions starting at Nash equilibrium policies, using (1), (2), (6), and (5).

We define *stable policies under* R to be policies $(\Lambda^R, \Lambda^{*R})$ such that neither nation can gain from reciprocal policy changes. Formally, for $(\Lambda_i, \Lambda_j) \in \{\Lambda \cup \Lambda^*\}^2$ evaluated at $(\Lambda^R, \Lambda^{*R})$,

$$\begin{aligned} \frac{dW}{d\Lambda_i} + \frac{dW}{d\Lambda_j}\frac{d\Lambda_j}{d\Lambda_i} &= 0, \text{ and} \\ \frac{dW^*}{d\Lambda_i} + \frac{dW^*}{d\Lambda_j}\frac{d\Lambda_j}{d\Lambda_i} &= 0. \end{aligned} \tag{8}$$

If $(\Lambda^R, \Lambda^{*R})$ exists, we can easily establish efficiency. Equations (8) imply policies are stable under R if for any policy pair $(\Lambda_i, \Lambda_j) \in \{\Lambda \cup \Lambda^*\}^2$, $\frac{dW}{d\Lambda_i} \neq 0$, $\frac{dW^*}{d\Lambda_i} \neq 0$, and

$$-\frac{d\Lambda_i}{d\Lambda_j} = \frac{\frac{dW}{d\Lambda_j}}{\frac{dW}{d\Lambda_i}} = \frac{\frac{dW^*}{d\Lambda_j}}{\frac{dW^*}{d\Lambda_i}}, \tag{9}$$

which then imply the equality in the efficiency conditions from (4),

$$\frac{\frac{dW}{d\Lambda_i}}{\frac{dW^*}{d\Lambda_i}} = \frac{\frac{dW}{d\Lambda_j}}{\frac{dW^*}{d\Lambda_j}}.$$

Given this tangency condition, efficiency then follows if we can ensure that these ratios are always negative, i.e. there is no possibility of Pareto gains from a single policy change. This can be confirmed with additional structure (as in the next subsection). We then state the following proposition illustrating the link between stability under reciprocity and efficiency:

Proposition 1 *Suppose* $(\Lambda^R, \Lambda^{*R})$ *is stable under reciprocity rule* R. *If for all* $\Lambda_i \in \{\Lambda \cup \Lambda^*\}$, $\frac{dW}{d\Lambda_i}$ *and* $\frac{dW^*}{d\Lambda_i}$ *have different sign (and are nonzero), then* $(\Lambda^R, \Lambda^{*R})$ *is on the global efficiency frontier.*

A relationship between stability under reciprocity and efficiency also exists in Bagwell and Staiger (1999, 2016). They derive a political optimum with desirable properties and show that it is stable under a particular reciprocity rule and efficient. What we have just shown is how stability under reciprocity can in fact imply efficiency. What we ultimately seek to establish is that this stability can be a more generally useful concept for directly selecting among possible policies on the efficiency frontier—even when the Bagwell and Staiger political optimum is not efficient.

2.3 Deriving Stable Points Under Reciprocity

A potential obstacle to applying stability under reciprocity to select an efficient point is that a stable point might not exist. At first glance, stability seems like a more demanding concept to satisfy than efficiency, as efficiency requires $|\Lambda| + |\Lambda^*|$ first-order conditions, while stability requires $2|\Lambda||\Lambda^*|$ conditions for 2 countries over all policy pairs. This conflict can be resolved by limiting ourselves to reciprocity rules that impose binding constraints on unilateral optimization. We show here that when nations are maximizing subject to a binding reciprocity rule, stability under reciprocity and efficiency are equivalent.

Define $\hat{W}(\Lambda^*; \Lambda_0, \Lambda_0^*)$ as the Home welfare achieved from new Foreign policy Λ^* given prevailing policies $(\Lambda^0, \Lambda^{0*})$ under R, such that

$$\hat{W}(\Lambda^*; \Lambda^0, \Lambda^{0*}) = \max_{\Lambda} W(\Lambda, \Lambda^*) \text{ subject to } R(\Lambda, \Lambda^*; \Lambda^0, \Lambda^{0*}) \geq 0. \quad \text{(H)}$$

We henceforth assume that R is binding. Let $\lambda^R(\Lambda^*; \Lambda^0, \Lambda^{0*}) > 0$ be the Lagrange multiplier at the solution to (H). We similarly define $\hat{W}^*(\Lambda, \Lambda^0, \Lambda^{0*})$

$$\hat{W}^*(\Lambda; \Lambda^0, \Lambda^{0*}) = \max_{\Lambda^*} W^*(\Lambda, \Lambda^*) \text{ subject to } R(\Lambda, \Lambda^*; \Lambda^0, \Lambda^{0*}) \leq 0. \quad \text{(F)}$$

and let $\lambda^{*R}(\Lambda; \Lambda_0, \Lambda_0^*) > 0$ be the multiplier at the solution to (F).

Notice that the derivatives of $\hat{W}$ and $\hat{W}^*$ can be interpreted as the gains in welfare from a small change in the trading partner's new policy. We then establish the following link between these problems and stability under reciprocity

Proposition 2 *For policies* $(\tilde{\Lambda}, \tilde{\Lambda}^*)$*, if* $\frac{d\hat{W}}{d\Lambda^*} = 0$ *evaluated at* $(\tilde{\Lambda}^*, \tilde{\Lambda}, \tilde{\Lambda}^*)$ *and* $\frac{d\hat{W}^*}{d\Lambda} = 0$ *evaluated at* $(\tilde{\Lambda}, \tilde{\Lambda}, \tilde{\Lambda}^*)$*, then* $(\tilde{\Lambda}, \tilde{\Lambda}^*)$ *are stable with respect to the reciprocity rule R. Furthermore,* $(\tilde{\Lambda}, \tilde{\Lambda}^*)$ *are the efficient policies* $(\hat{\Lambda}(\hat{\lambda}^R), \hat{\Lambda}^*(\hat{\lambda}^R))$ *with* $\hat{\lambda}^R \equiv \tilde{\lambda}^R / \tilde{\lambda}^{R*}$*,* $\tilde{\lambda}^R \equiv \lambda^R(\tilde{\Lambda}^*, \tilde{\Lambda}, \tilde{\Lambda}^*)$*, and* $\tilde{\lambda}^R \equiv \lambda^{*R}(\tilde{\Lambda}, \tilde{\Lambda}^*, \tilde{\Lambda}^*)$*.*

Proof. The equalities $\frac{d\hat{W}}{d\Lambda^*} = 0$ and $\frac{d\hat{W}^*}{d\Lambda} = 0$ imply that $(\tilde{\Lambda}, \tilde{\Lambda}^*)$ satisfy the following $|\Lambda| + |\Lambda^*|$ first-order conditions derived from the envelope theorem

$$\begin{aligned} \frac{dW}{d\Lambda^*} + \tilde{\lambda}^R \frac{dR}{d\Lambda^*} &= 0, \text{ and} \\ \frac{dW^*}{d\Lambda} - \tilde{\lambda}^{R*} \frac{dR}{d\Lambda} &= 0. \end{aligned} \tag{10}$$

Since the constraint in R binds at the solution and R is monotonic, we must have that $\tilde{\Lambda}$ is the argmax of the optimization problem in (H) and $\tilde{\Lambda}^*$ is the argmax of the optimization problem in (F). From optimization of (H) and (F), the following also holds at $(\tilde{\Lambda}, \tilde{\Lambda}^*)$

$$\begin{aligned} \frac{dW}{d\Lambda} + \tilde{\lambda}^R \frac{dR}{d\Lambda} &= 0, \text{ and} \\ \frac{dW^*}{d\Lambda^*} - \tilde{\lambda}^{R*} \frac{dR}{d\Lambda^*} &= 0. \end{aligned} \tag{11}$$

Equations (10) and (11) imply that for any $\Lambda_i \in \Lambda \cup \Lambda^*$,

$$\frac{\frac{dW}{d\Lambda_i}}{\frac{dW^*}{d\Lambda_i}} = -\frac{\tilde{\lambda}^R}{\tilde{\lambda}^{R*}} < 0.$$

The efficiency conditions (4) are then satisfied and $(\tilde{\Lambda}, \tilde{\Lambda}^*) \in EF$. Lastly, consider stability under reciprocity. Take arbitrary $(\Lambda_i, \Lambda_j) \in \{\Lambda \cup \Lambda^*\}^2$. If we combine pairs of the individual policy derivatives from either (10) or (11) by substituting out $\tilde{\lambda}^R$ or $\tilde{\lambda}^{R*}$ as appropriate, and substitute in $\frac{d\Lambda_j}{d\Lambda_i}$ using its definition from (5), then each equation (8) that defines stable policies under R is satisfied. ■

The approach here is a generalization of Bagwell and Staiger (2001a), who show that efficient trade and domestic policies can be achieved by subjecting unilateral choices to a market access preservation constraint, and this constraint can take the form of a reciprocity rule R above.[5] The proposition establishes that such an approach can be used more broadly to determine policies that are both efficient and stable under reciprocity. The approach will

[5]See problems (IV) and (IV*) in Bagwell and Staiger (2001a). The constraints there are defined to preserve a trading partner's market access. Their constraints can be expressed in terms of a common function R by multiplying by the appropriate relative world prices and applying balanced trade conditions.

be helpful as we consider a specific reciprocity rule, and apply it to deriving predictions when trade agreements have a role in addressing local price externalities.

3 Generalizing Standard Reciprocity Results

This section adopts a specific reciprocity rule, which we call the standard reciprocity rule. The rule ensures that policy changes increase the value of trade equally at prevailing world prices. We focus on this rule throughout the rest of the paper, because this rule has already been shown to be consistent with efficient outcomes when trade agreements need only eliminate terms-of-trade manipulation (Bagwell and Staiger, 2016). Our approach is to first generalize this result as much as possible, before evaluating in the next section whether the same rule can apply when trade agreements address local price externalities.

Though we focus here on the standard reciprocity rule, there is one established alternative—rules that ensure that trade-expanding policy changes preserve sectoral trade balances. Such a description applies to reciprocity rules proposed in intra-industry trade models by Ossa (2011) and Mrázová (2011). Each model is partial equilibrium and preserves trade balance in both an outside sector and an imperfectly competitive sector. In certain situations, the sectoral-balance rule and the standard rule are in fact equivalent. The rule in Ossa (2011) is equivalent to a standard reciprocity rule when one defines the world price to be the world price index of traded goods.[6] Mrázová shows her rule is equivalent to the standard reciprocity rule when nations have linear demands, though this equivalence fails when demands are nonlinear. We discuss the possibility of alternative rules further in the conclusion.

We proceed as follows. First, we add more structure so policies affect government objectives only through prices. We then define the standard reciprocity rule in the context of this model. We define policy space completeness, such that nations can achieve policies where there are no first-order effects from local price changes—as in the Bagwell and Staiger political optimum—and we prove these policies are stable and efficient. We then categorize existing literature based on how policy space completeness is achieved, and when it is known to fail. By generalizing in this section when local price externalities do not matter, we clarify the theoretical settings in which local price externalities do matter, and where there is value in developing new predictions.

[6] In Ossa's model, policies that are reciprocal also preserve the number of firms, and local prices of domestic goods are constant, so reciprocal policy changes also preserve the world price index. DeRemer (2013b), Bergin and Corsetti (2013), Costinot, Rodriguez-Clare, and Werning (2016) all support the claim that the price index of the traded bundle is an appropriate definition of terms of trade in such differentiated product settings. One difference from the standard rule though is that unilateral trade-expanding policies actually worsen the terms of trade in Ossa's model, but still increase the total value of exports.

3.1 A Trade Model with Local and World Prices

This section specifies a particular trade model that generalizes Bagwell and Staiger (1999) to many different policies and the possibility that there are externalities from local prices, in addition to terms-of-trade externalities. Consider two countries, Home and Foreign, with reduced-form government objectives $W(p^l(\Lambda, \Lambda^*), p^w(\Lambda, \Lambda^*))$ for Home and $W^*(p^{l*}(\Lambda, \Lambda^*), p^w(\Lambda, \Lambda^*))$ for Foreign, which satisfy the conditions specified in Section 2.1. The vectors p^l and p^{l*} contain local prices relevant for objectives in either Home or Foreign's market. So if p is the vector of local prices in Home's market (of either Home or Foreign-produced goods) and p^* is similarly the vector of local prices in Foreign's market, then p^l, $p^{l*} \subseteq p \cup p^*$. The vector p^w contains all world prices between borders that then constitute the terms of trade. The local price vectors could include prices for factors of production, intermediate goods, final goods, or composite goods. Let there be at least one factor of production that is fully mobile between sectors in each country, and designate the Home wage for this factor as the numeraire.

The row matrices for the effect of one's own policies on one's own objective are then

$$\frac{dW}{d\Lambda} = \frac{dW}{dp^l}\frac{dp^l}{d\Lambda} + \frac{dW}{dp^w}\frac{dp^w}{d\Lambda}, \text{ and } \frac{dW^*}{d\Lambda^*} = \frac{dW^*}{dp^{l*}}\frac{dp^{l*}}{d\Lambda^*} + \frac{dW^*}{dp^w}\frac{dp^w}{d\Lambda^*}, \tag{12}$$

and the effects of trading partner's policies on one's own objective are

$$\frac{dW}{d\Lambda^*} = \frac{dW}{dp^l}\frac{dp^l}{d\Lambda^*} + \frac{dW}{dp^w}\frac{dp^w}{d\Lambda^*}, \text{ and } \frac{dW^*}{d\Lambda} = \frac{dW^*}{dp^{l*}}\frac{dp^{l*}}{d\Lambda} + \frac{dW^*}{dp^w}\frac{dp^w}{d\Lambda}. \tag{13}$$

We make several assumptions about trade. We assume throughout that a *balanced trade condition* is satisfied for all policy choices. Let M be a row vector of net trade volumes for each traded item, in which export items for Home enter positively and import items for Home enter negatively. We can then write the balanced trade condition as $Mp^W = 0$.

Without loss of generality, we assume a nation's own policies are defined so that a nation's own policies are trade-weighted terms-of-trade improving at the Nash equilibrium:

$$M\frac{dp^w}{d\Lambda} > 0 \text{ and } M\frac{dp^w}{d\Lambda^*} < 0. \tag{14}$$

This assumption implies that for import goods, nations prefer a lower world price, and for export goods, nations prefer a higher world price, all else equal.[7]

One final assumption is that the policy vectors include import tariffs, which we denote

[7] Throughout we consider M based on net trade flows between locations, but the theory and definition of M could also be modified to consider governments that value offshore production owned by multinationals (e.g. Blanchard, 2010).

as $\Lambda^\tau \subseteq \Lambda$ and $\Lambda^{\tau *} \subseteq \Lambda^*$, which drive a wedge between world prices and a subset of import market local prices that we denote as $p^{l\tau}$ and $p^{l\tau *}$. These tariffs improve the terms of trade for importing countries and increase local prices (i.e. no Metzler or Lerner paradoxes). So formally,

$$\frac{dp^{l\tau}}{d\Lambda^\tau} > 0 > \frac{dp^w}{d\Lambda^\tau} \text{ and } \frac{dp^{l\tau *}}{d\Lambda^{\tau *}} > 0 > \frac{dp^{w*}}{d\Lambda^{\tau *}}. \tag{15}$$

And lastly, import volumes are decreasing in the local price.

$$\frac{dM}{dp^{l\tau}}\frac{dp^{l\tau}}{d\Lambda^\tau} > 0 \text{ and } \frac{dM}{dp^{l\tau *}}\frac{dp^{l\tau *}}{d\Lambda^{\tau *}} < 0. \tag{16}$$

For Home, the import volume is defined negatively, so this derivative is positive in our notation.

We further assume that the effects of world prices on objectives are proportional to the trade volumes, with the same sign. We denote these factors as $\frac{dW}{dT} > 0$ and $\frac{dW^*}{dT^*} < 0$, which are both strictly positive. We then have

$$\frac{dW}{dp^w} = \frac{dW}{dT}M \text{ and } \frac{dW^*}{dp^w} = \frac{dW^*}{dT^*}M. \tag{17}$$

The restriction permits any model where the effect of terms of trade on welfare occurs via total trade tax revenue (e.g. Bagwell and Staiger, 2005a) or if trade is between only two goods (e.g. Bagwell and Staiger, 1999). This is the last assumption we consider here. The setting is general enough to encompass most static two-country models where the role of trade agreements is to address international externalities.[8]

Before moving on to consider the standard reciprocity rule, we also consider price derivative properties that are important for interpretation. First, consider general effects through the world price derivatives. Combining (14) and (17), own policy increases through terms of trade are positive for nations imposing them and negative for the trading partners

$$\begin{aligned} \frac{dW}{dp^w}\frac{dp^w}{d\Lambda} &> 0, \frac{dW}{dp^w}\frac{dp^w}{d\Lambda^*} < 0, \\ \frac{dW^*}{dp^w}\frac{dp^w}{d\Lambda^*} &> 0, \text{ and } \frac{dW^*}{dp^w}\frac{dp^w}{d\Lambda} < 0. \end{aligned} \tag{18}$$

Notice that these inequalities almost imply our assumption (2) that $\frac{dW}{d\Lambda^*} < 0$ and $\frac{dW^*}{d\Lambda} < 0$

[8] We focus on interior solutions and cannot encompass models in which the equilibria are corner solutions, such as in Ossa (2011) or DeRemer (forthcoming), where import subsidies are exogenously ruled out and the cooperative equilibrium is free trade. The current framework can encompass a model similar to Ossa with endogenous choice between import tariffs and subsidies, such as Campolmi, Fadinger, and Forlati (2014).

when evaluated at Nash equilibrium policies, but local price effects could also affect $\frac{dW}{d\Lambda^*}$ and $\frac{dW^*}{d\Lambda}$. We allow such local price externalities to exist with either sign. If they mute rather than amplify the negative terms-of-trade effects, then for our assumptions to be consistent, the local price effects are not large enough to dominate, so $\frac{dW}{d\Lambda^*} < 0$ and $\frac{dW^*}{d\Lambda} < 0$ would still hold at the Nash equilibrium.

As for effects through local prices, (1), (12) and (18) imply that at Nash policies

$$\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} < 0 \text{ and } \frac{dW^*}{dp^{l*}}\frac{dp^{l*}}{d\Lambda^*} < 0. \tag{19}$$

For prices of imported products, (15) and (19) imply that

$$\frac{dW}{dp^{l\tau}} < 0 \text{ and } \frac{dW}{dp^{l\tau*}} < 0. \tag{20}$$

As is standard, the Nash equilibrium is characterized by policies such that nations would benefit from the (import-competing sector) local price decrease that would come with an import tariff reduction, but they choose not to reduce the tariff unilaterally because the same reduction would also cause a terms-of-trade loss.

3.2 The Standard Reciprocity Rule

We introduce a generalization of the standard reciprocity rule (e.g. Bagwell and Staiger, 1999, 2016) such that policy changes affect trade volumes equally when valued at world prices. Define policies (Λ, Λ^*) to satisfy *the standard reciprocity rule* R^0 at $(\Lambda^0, \Lambda^{*0})$ if

$$R^0(\Lambda, \Lambda^*; \Lambda^0, \Lambda^{*0}) = M(\Lambda, \Lambda^*)p^w(\Lambda^0, \Lambda^{*0}) = 0.$$

We establish multiple economic interpretations for this rule. When totally differentiating the rule, the resulting equation makes transparent the interpretation that reciprocal policy changes must affect trade equally when valued at world prices:

$$dR^0 = \left(\frac{dM}{d\Lambda}d\Lambda + \frac{dM}{d\Lambda^*}d\Lambda^*\right)^T p^w = 0. \tag{21}$$

Subtracting the total derivative of the balanced trade condition, we have

$$dR^0 = -M\left(\frac{dp^w}{d\Lambda}d\Lambda + \frac{dp^w}{d\Lambda^*}d\Lambda^*\right) = 0. \tag{22}$$

This last condition is satisfied if policies are neutral with respect to world prices. They also can be satisfied if policy changes are trade tax neutral. To determine $\frac{d\Lambda_j}{d\Lambda_i}$ for any policy pairs

$(\Lambda_i, \Lambda_j) \in \{\Lambda \cup \Lambda^*\}^2$, the above conditions imply the following representations

$$-\frac{d\Lambda_j}{d\Lambda_i} = \frac{\frac{dM}{d\Lambda_i}^T p^w}{\frac{dM}{d\Lambda_j}^T p^w} = \frac{M\frac{dp^w}{d\Lambda_i}}{M\frac{dp^w}{d\Lambda_j}}. \tag{23}$$

So reciprocal policy changes increase net trade volume equally at prevailing world prices and keep the trade-weighted terms of trade constant.

Next we confirm that the standard reciprocity rule satisfies the conditions of the generic reciprocity rule in Section 2.2, so we can apply our general results. We find that $\frac{dR^0}{d\Lambda} = -M\frac{dp^w}{d\Lambda} < 0$ and $\frac{dR^0}{d\Lambda^*} = -M\frac{dp^w}{d\Lambda^*} > 0$, using (14) and (22). So we can interpret the rule as implying constraints $R \geq 0$ for Home and $R \leq 0$ for Foreign, and we have $\frac{d\Lambda_j^*}{d\Lambda_i} > 0$ for $(\Lambda_i, \Lambda_j^*) \in \Lambda \times \Lambda^*$. The constraints prevent either nation from policy changes that improve terms of trade, and they must be binding whenever policies are constrained away from the Nash equilibrium. Having confirmed these properties, we can apply results of Section 2.2 and 2.3.

Remark 1 *Stable policies under the standard reciprocity rule R^0 must be efficient, and nations must gain from reciprocal policy reductions starting from the Nash equilibrium.*

Because of the additional structure we have in the current model, we can further develop properties concerning what we will call the *total local price effects of reciprocal policy reductions.* Using (7), (12) and (22), gains from reciprocal policy reduction are possible if

$$\begin{aligned} \frac{dW}{dp^l}\left(\frac{dp^l}{d\Lambda} + \frac{dp^l}{d\Lambda^*}\frac{d\Lambda^*}{d\Lambda}\right) &< 0, \text{ and} \\ \frac{dW^*}{dp^{l*}}\left(\frac{dp^{l*}}{d\Lambda} + \frac{dp^{l*}}{d\Lambda^*}\frac{d\Lambda^*}{d\Lambda}\right) &< 0. \end{aligned} \tag{24}$$

Then using (8), (12) and (22), we find that at the stable and efficient policies

$$\begin{aligned} \frac{dW}{dp^l}\left(\frac{dp^l}{d\Lambda} + \frac{dp^l}{d\Lambda^*}\frac{d\Lambda^*}{d\Lambda}\right) &= 0, \text{ and} \\ \frac{dW^*}{dp^{l*}}\left(\frac{dp^{l*}}{d\Lambda} + \frac{dp^{l*}}{d\Lambda^*}\frac{d\Lambda^*}{d\Lambda}\right) &= 0. \end{aligned} \tag{25}$$

The progression from Nash policies to the efficient point then represents a transition from negative total local price effects of reciprocal policy changes, to zero total local price effects of reciprocal policy changes. From (15) and (16) we can conclude that this progress can also be interpreted as an increase in import volumes from low Nash levels to higher cooperative levels.

Notice how this contrasts with previous theory of reciprocity (Bagwell and Staiger, 2016) in our notation. There the progression to efficiency occurs through gains from one's own policy reductions $\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} < 0$ until stability is achieved at $\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} = 0$.

Importantly, unlike the previous theory with import tariffs, we can no longer conclude that $\frac{dW}{dp^l} = 0$ and $\frac{dW}{dp^{l*}} = 0$ must hold at the stable and efficient point when local price externalities matter. This leaves open the new possibility of having a stable equilibrium where these derivatives are positive, and each nation seeks price increases in import-competing sectors. We discuss the implications of this possibility further in Section 4.

3.3 Efficiency When Policies Are Complete

Having established the standard reciprocity rule, we introduce a set of policies that has been proven to be efficient and stable in prior literature. We seek to extend this result as much as possible, so we know better when we need an alternative approach to find the stable and efficient point. We re-introduce the *Bagwell and Staiger political optimum*, a set of policies $(\Lambda_{PO}, \Lambda^*_{PO})$ that nations choose unilaterally if they act as if they do not value the ability to manipulate the terms of trade p^w. Using (12), the following conditions hold at $(\Lambda_{PO}, \Lambda^*_{PO})$:

$$\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} = \frac{dW^*}{dp^{l*}}\frac{dp^{l*}}{d\Lambda^*} = 0. \tag{26}$$

So local price effects are zero at the political optimum, rather than negative like at the Nash equilibrium. Bagwell and Staiger (2016) survey many circumstances such that their political optimum lies on the efficiency frontier, even if W and W^* reflect political motives, hence the terminology.

What we now show is that the political optimum is stable and efficient if the policy space exhibits particular forms of completeness. We derive two propositions to this effect. The first proposition considers policy space completeness as when nations have the capacity to achieve an efficient allocation of production q according to their objectives.

Proposition 3 *Suppose welfare can also be written as functions $W(q(p, p^*, p^w), p^w)$ and $W^*(q(p, p^*, p^w), p^w)$, but still consistent with Section 3.1. Suppose there exist (first-best) policies $(\Lambda_{FB}, \Lambda^*_{FB})$ such that $\frac{dW}{dq} = 0$ and $\frac{dW^*}{dq} = 0$. Then these policies are efficient, and stable only with respect to the standard reciprocity rule R^0.*

Proof. At $(\Lambda_{FB}, \Lambda^*_{FB})$, all cross-border policy effects through local prices must be zero, so for any $\Lambda_i \in \Lambda \cup \Lambda^*$,

$$\frac{dW}{d\Lambda_i} = \frac{\partial W}{\partial p^w}\frac{dp^W}{d\Lambda_i} \text{ and } \frac{dW^*}{d\Lambda_i} = \frac{\partial W^*}{\partial p^w}\frac{dp^W}{d\Lambda_i}.$$

Then from assumption (17) and (23), for any $(\Lambda_i, \Lambda_j) \in (\Lambda \cup \Lambda^*)^2$,

$$\frac{\frac{dW}{d\Lambda_i}}{\frac{dW}{d\Lambda_j}} = \frac{\frac{dW^*}{d\Lambda_i}}{\frac{dW^*}{d\Lambda_j}} = \frac{M\frac{dp^w}{d\Lambda_i}}{M\frac{dp^w}{d\Lambda_j}} = -\frac{d\Lambda_i}{d\Lambda_j}.$$

So under the standard reciprocity rule, (9) is satisfied, and so $(\Lambda_{FB}, \Lambda^*_{FB})$ are efficient and stable under the standard reciprocity rule in (23). ■

The proposition says that if nations have enough policy space to achieve a first-best allocation, then terms-of-trade manipulation is the only problem for trade agreements to solve. Must local price externalities then matter if the first-best allocation cannot be achieved? The following proposition establishes that the answer is no, as another form of policy completeness is still attainable.

Proposition 4 *Suppose the Bagwell and Staiger political optimum $(\Lambda_{PO}, \Lambda^*_{PO})$ exists, and $\frac{dp^l}{d\Lambda}$ and $\frac{dp^{l*}}{d\Lambda^*}$ have full row rank. Then these policies are efficient, and stable only with respect to the standard reciprocity rule R^0.*

Proof. Because $\frac{dp^l}{d\Lambda}$ and $\frac{dp^{l*}}{d\Lambda^*}$ have full row rank, their cokernel contains only the 0 vector. So the conditions for the political optimum (26) imply that $\frac{dW}{dp^l} = 0$ and $\frac{dW*}{dp^{l*}} = 0$. By (12) and (13), all cross-border policy effects through local prices must be zero. The rest of the proof then proceeds by the same argument as the previous proposition. ■

The proposition implies that if nations' policy space is sufficient for each to independently affect all local prices relevant for its objectives, whether the prices are in the domestic market or abroad, then terms-of-trade externalities are the only international externalities that trade agreements need to address.

A natural question is then whether the failure of the two propositions is sufficient to establish that local price externalities matter for trade agreements. If Proposition 4 fails, then (without loss of generality) there exists some $p^L \subseteq p^l$ such that Home cannot independently influence the local price, so we cannot determine whether or not $\frac{dW}{dp^L}$ is zero. Sufficiency then requires additional structure that would allow us to determine that this derivative vector is indeed nonzero.

3.4 When Do Local Price Externalities Matter, or Not?

We have just derived two propositions that detail when local price externalities do not matter for trade agreements. Building on prior literature, we discuss which examples satisfy the conditions for either of the two propositions, and which examples fail and leave a potential role for trade agreements to address local price externalities.

First, notice that Proposition 3 is akin to the targeting principle (Bhagwati and Ramaswami, 1963; Dixit, 1985). If governments have enough domestic policies to address any distortion related to local prices (whether the distortion derives from imperfect competition, externalities, or its own political preferences), then there is no need to distort international prices to address problems with local prices. When the allocation is efficient, there is no first-order gain from any policies that shift local prices.

Proposition 4 is a generalization of results that were first emphasized by Bagwell and Staiger (2012b, 2015). Having a complete policy space of both import and export policies implies that their political optimum is efficient, and local price externalities are no longer relevant for trade agreements, even if the conditions for Proposition 3 fail. The Proposition generalizes their results, as they consider only the case where import and export policies are perfect substitutes in determining trade costs, which then determine the local prices of traded goods.[9] Proposition 4 also generalizes their results to cases with domestic policies.[10]

We can illustrate either the satisfaction or failure of the conditions for the two propositions using models of imperfect competition with an arbitrary number of sectors. If there is no intersectoral misallocation based on sectoral differences in markups—either because there is only a single sector, or the market outcome ensures constant markups across sectors, or governments have subsidies to equate markups across sectors—then there are no international externalities related to imperfect competition (see e.g. Epifani and Gancia, 2011). Imperfect competition only starts to matter once such subsidies to equate markups across sectors are limited, and nations also lack the complete trade policy space of import and export policies as in Proposition 4 (as in Ossa, 2011).

So if nations have sufficient policy space to affect all prices, local price externalities are irrelevant. Yet governments have long turned to trade policy to address fundamentally

[9] Bagwell and Lee (2015) also find in their model that the political optimum is efficient, even though the import and export policies are not perfect substitutes.

[10] Proposition 4 confirms stability under reciprocity rule R^0 even when intermediate prices are determined under bilateral bargaining, as in the offshoring model of Antràs and Staiger (2012a), who consider both domestic policies and trade policies. Stability under reciprocity may seem puzzling in their model given footnote 5 and their finding that market access preservation does not ensure efficiency. The distinction is that they consider preservation only in their intermediate trade volume, and argue that such preservation does not prevent terms-of-trade manipulation. Bilateral bargaining does indeed break the usual link between preservation of trade volume for one sector and preventing terms-of-trade manipulation. But if we instead define market access preservation in Antràs and Staiger based on the total value of trade at prevailing world prices, as in equation (21), then this total value also includes larger trade volume in the outside sector when a nation manipulates its policy mix to improve its terms of trade. Then the usual argument still applies that trade value preservation and balanced trade imply terms-of-trade preservation, as in equation (22), so this broader form of market access preservation still ensures efficiency.

domestic issues. For example, Alexander Hamilton's 1791 proposal for promoting U.S. manufacturing involved significant "bounties" (e.g. domestic subsidies) but the United States instead turned to trade policy to support manufacturing (Irwin, 2004). Particular models consider costs of administering subsidies versus tariffs (e.g. Ederington and Minier, 2008). But there is no consensus answer for why nations do not subsidize from the political economy of trade policy, according to surveys by Rodrik (1995) and McLaren (2016). Restricting subsidies exogenously and studying the implications then remains a common approach.

We now preview why the examples we consider in Section 5 fail to satisfy the two propositions. The first case is a three-sector, perfect competition, partial equilibrium model of Bagwell and Staiger (2001b), which Bagwell and Staiger (2016) revisit with limited instruments. When nations have both import and export policies in this setting, the conditions for Proposition 4 are met. Absent export subsidies and domestic subsidies, the government has no way to affect the price of its export industry, but the conditions for Proposition 4 are still met when there are no political economy forces affecting the export sector. In this case, there are no rents to be gained through higher output, and therefore no local price externalities. But if political economy affects the export sector and neither domestic nor export subsidies are available, then neither of the two propositions applies. As confirmed by Bagwell and Staiger (2016), the political optimum is inefficient, and we can then show in the current paper how a trade agreement can play a role in addressing the local price externality.

The offshoring model of Antràs and Staiger (2012b) is another in which the authors establish that the political optimum is inefficient, so there is a role for trade agreements to address local price externalities. For governments to lack the policy space to achieve the first-best here, they require both political economy and intermediate prices determined by bilateral bargaining. With no political economy, there is enough policy space for the price of the intermediate export to be set equal its marginal cost, and then the intermediate exporter has no interest in its trade volume, so Proposition 3 applies. With no bilateral bargaining, only the prices of the intermediate matter for determining trade volume, not the price of the final good, and again governments can achieve an efficient allocation based on Proposition 3. But with both political economy and bilateral bargaining, governments distort the price of the intermediate above its marginal cost, and the final good price matters for determining intermediate trade volume. There is then no dual instrument for the final good tariff such that the conditions of Proposition 4 are satisfied. The authors then confirm that there is still a local price externality that matters for the intermediate exporter's nation in trade negotiations.

4 Reciprocity with Local Price Externalities

This section considers the standard reciprocity rule when trade agreements have a role in addressing local price externalities. The goal is to determine whether the reciprocity rule can select an efficient trade agreement outcome when the Bagwell and Staiger political optimum is inefficient. We then explore distinct features of the new outcome.

The approach is not as straightforward as deriving the usual stable outcome under reciprocity, because the equilibrium does not always exist. We first determine general properties of stable outcomes under reciprocity, given existence, so we can then determine some sufficient conditions for when existence fails. Guided by these results, we develop an approach to find stability under a limited policy space. We show that reciprocity can still yield a Pareto efficient result with respect to the full policy space, and we argue that such a limited policy space is still broadly consistent with the structure of actual GATT/WTO negotiations. We then discuss properties of the stable equilibrium, focusing on the possibility that governments can benefit from local price increases in import-competing sectors.

4.1 General Properties of Stable Policies Under Reciprocity

We first derive general properties of stable policies under the standard reciprocity rule. The first proposition states that when policies are stable under reciprocity, the derivatives of local price effects must have a common sign (or all equal zero).

Proposition 5 *If policies* $(\Lambda^{R^0}, \Lambda^{*R^0})$ *are stable under reciprocity rule* R^0*, then all elements of* $\left\{\frac{dW}{dp^l}\frac{dp^l}{d\Lambda_i} \mid \Lambda_i \in \Lambda \cup \Lambda^*\right\}$ *must all have the same sign or all equal zero. The same result holds for* $\left\{\frac{dW^*}{dp^{l*}}\frac{dp^{l*}}{d\Lambda_i} \mid \Lambda_i \in \Lambda \cup \Lambda^*\right\}$.

Proof. Consider arbitrary $(\Lambda_i, \Lambda_j) \in \{\Lambda \cup \Lambda^*\}^2$. By (9), (12), (17), and (23)

$$\frac{\frac{dW}{dp^l}\frac{dp^l}{d\Lambda_i} + \frac{dW}{dT}M\frac{dp^w}{d\Lambda_i}}{\frac{dW}{dp^l}\frac{dp^l}{d\Lambda_j} + \frac{dW}{dT}M\frac{dp^w}{d\Lambda_j}} = \frac{M\frac{dp^w}{d\Lambda_i}}{M\frac{dp^w}{d\Lambda_j}},$$

Then straightforward algebra yields

$$\frac{\frac{dW}{dp^l}\frac{dp^l}{d\Lambda_i}}{M\frac{dp^w}{d\Lambda_i}} = \frac{\frac{dW}{dp^l}\frac{dp^l}{d\Lambda_j}}{M\frac{dp^w}{d\Lambda_j}}.$$

As we have defined policies so the denominators are positive, the proposition must hold for Home welfare W and prices p_l. The same argument then holds for Foreign welfare W^* and prices p_l^*. ■

The proposition implies that if policies have the same relative effect on total welfare as on the trade-weighted terms of trade, then policies must have the same relative effect through local prices (or no effect). Consequently, the local price effects must either be the same sign or zero.

The result makes apparent an obstacle for using the standard reciprocity rule as a guide to the efficiency frontier. Suppose that the local price externalities are persistent regardless of policy. Then the proposition implies that if governments have sufficient policy space to address local price externalities in some sectors, then the stable outcome under reciprocity must eliminate local price externalities in all sectors, so there is no stable outcome.

For example, if nations have an import tax and export tax affecting the same local price, then stability under reciprocity implies this local price derivative is zero for both nations—otherwise one nation would gain from a terms-of-trade preserving policy change that changes the price. If there is then some persistent local price externality, then the equilibrium cannot exist. Bagwell and Staiger (2016) observe that this is indeed the case in Antràs and Staiger (2012b), who focus on a case in which Foreign supplies an intermediate good to Home, Home and Foreign each have trade policies affecting trade in the intermediate, and Home has a policy affecting the final good price. The stable outcome then must involve zero local price effects from the intermediate prices on welfare, and then there is a persistent local price externality from the final good price. By Proposition 5, a stable outcome cannot exist in this case. The following proposition, proven in the appendix, generalizes the existence failure found in Antràs and Staiger:

Proposition 6 *Suppose that for any policies, there is some $p_i \in p^l \cap p^{l*}$, such that each nation can influence p_i independently from other prices, and p_i is not fully determined by p^w. Then if $(\Lambda^{R^0}, \Lambda^{*R^0})$ exists, it is equivalent to $(\Lambda^{PO}, \Lambda^{*PO})$. If $(\Lambda^{PO}, \Lambda^{*PO}) \notin EF$, then no stable point under the standard reciprocity rule exists.*

The proposition confirms that reciprocity over the full policy space cannot always guide nations to the efficiency frontier. The problem is not when a nation has redundancy in both trade and domestic policies that influence the same local price, but when multiple nations can independently influence the same local price. This proposition then motivates an approach where we seek to reduce the number of policies under consideration in reciprocal negotiations in order to facilitate existence, but without compromising the possibility of reaching the efficiency frontier.

4.2 Limited Reciprocity When Local Price Externalities Matter

When the stable outcome under reciprocity does not exist, a promising approach is to limit the scope of negotiations under reciprocity to a subset of policies such that a stable outcome could exist. This approach is also broadly consistent with actual GATT/WTO norms and rules. In traditional trade negotiations, we observe reciprocity over import tariffs, not both import and export policies.

A reasonable concern though is whether limiting the policy space for reciprocal negotiations would lead to a Pareto inferior outcome, but the following Proposition establishes that such limitation need not be an obstacle to achieving efficiency.

Proposition 7 *Suppose* $\bar{\Lambda} \subseteq \Lambda$ *and* $\bar{\Lambda}^* \subseteq \Lambda^*$ *are constrained policy spaces that have the same image as* (Λ, Λ^*) *with respect to the functions* $p^l(\Lambda, \Lambda^*)$ *and* $p^{l*}(\Lambda, \Lambda^*)$. *Then* $EF(\bar{\Lambda}, \bar{\Lambda}^*) \subseteq EF(\Lambda, \Lambda^*)$. *So if* $(\bar{\Lambda}^{R^0}, \bar{\Lambda}^{*R^0})$ *is stable under* R^0 *within the domain* $(\bar{\Lambda}, \bar{\Lambda}^*)$, *then* $(\bar{\Lambda}^{R^0}, \bar{\Lambda}^{*R^0}) \in EF(\Lambda, \Lambda^*)$.

Proof. Consider an arbitrary point $(\Lambda^\lambda, \Lambda^{*\lambda}) \in EF(\Lambda, \Lambda^*)$. Then by assumption there exists $(\bar{\Lambda}^\lambda, \bar{\Lambda}^{*\lambda}) \in (\bar{\Lambda}, \bar{\Lambda}^*)$ such that both points imply the same p^l and p^{l*}. The difference in the two outcomes is then differences in p^w, but by (17), shifts solely in p^w cannot be Pareto-improving. Thus, $(\bar{\Lambda}^\lambda, \bar{\Lambda}^{*\lambda})$ cannot be Pareto inferior to $(\Lambda^\lambda, \Lambda^{*\lambda})$. Thus $EF(\bar{\Lambda}, \bar{\Lambda}^*) \subseteq EF(\Lambda, \Lambda^*)$. Then for $(\bar{\Lambda}^{R^0}, \bar{\Lambda}^{*R^0})$, stability under R^0 within $(\bar{\Lambda}, \bar{\Lambda}^*)$ implies (by Proposition 1) that $(\bar{\Lambda}^{R^0}, \bar{\Lambda}^{*R^0}) \in EF(\bar{\Lambda}, \bar{\Lambda}^*)$, which we have just shown is contained in $EF(\Lambda, \Lambda^*)$. ■

The Proposition implies that international redundancy in policies affecting local market prices—a redundancy that causes instability under reciprocity in Proposition 6—does not offer any Pareto improvement. The difference in outcomes offered by such redundancy is the ability to make zero-sum transfers through policy changes that preserve local prices but shift rents between nations via changes in the terms of trade. Limiting the policy space as described is then purely redistributional and does not imply any efficiency loss.

A recent literature considering reciprocity over both import and export policies (Bagwell and Staiger, 2012b, 2015) observes that a "missing instruments" problem from banning export policies complicates the problems for trade agreements to solve, and that trade agreements can then play a role in substituting for these missing instruments in addressing these additional externalities. For the stable point under the standard reciprocity rule over limited instruments, a trade agreement indeed prevents any loss in efficiency from limiting the instruments under consideration.

The proposition leaves open the question of exactly which instruments nations would select for negotiations. An approach broadly consistent with the GATT/WTO is that import policies are permitted and negotiated over, export policies are prohibited, and domestic policies are permitted as long as they do not undermine market access expectations implied by import policies, or in the case of domestic subsidies, cause serious prejudice to trading partners (see e.g. Bagwell, Bown, and Staiger, 2016). Such an approach then gives nations ability to set local prices as desired for import-competing sectors, while desired local prices for export industries must be achieved through trade agreements. This approach prevents the cross-country redundancy in policies that poses problems for the existence of stable policies under reciprocity.

Why would the GATT/WTO settle on the aforementioned approach, rather than one that focuses on prohibiting import policies?[11] We cannot answer this question within the current paper's framework, though we discuss a possible answer here consistent with prior trade agreement literature. One of the key tradeoffs, which is relatively neglected in the literature, is that such a choice involves not only prohibiting one type of border measure or the other, but also there needs to be deeper integration to prevent domestic policies that could substitute for the prohibited trade policies (as opposed to a shallow integration that merely prevents substitution for border measures that have been reduced gradually). So if the costs of deep integration for domestic policies that affect import-competing industries is larger than costs of deep integration for policies that affect export industries, then this motivates negotiating over import policies while prohibiting export policies. The existing literature argues that directly contracting over domestic policies affecting import-competing industries is costly (as in Horn, Maggi, and Staiger, 2010), and a shallow integration approach is preferable.[12] As Bagwell and Staiger (2001a) explain, the GATT did prevent governments from using domestic policies that would undermine market access anticipated from negotiated reciprocal tariff reductions, as codified in the GATT Article XXIII nonviolation complaint, and the theory finds support in GATT history and legal scholarship.[13] Later WTO agreements use a similar shallow integration approach for non-tariff measures, which countries can use to achieve legitimate regulatory goals as long as they are not more trade restrictive than necessary (Staiger, 2015). An exception is the deep integration of the WTO subsidy rules, which legislate that domestic subsidies to either import-competing or export industries can be removed if they cause adverse effects for trading partners (Sykes, 2005). On balance though,

[11] There is a large literature on why restrictions on export subsidies are prohibited (for surveys, see Lee, 2016; Bagwell, Bown, and Staiger, 2016). As all authors point out, the literature does not fully address the question of why one policy or the other is negotiated over gradually while the other is restricted.

[12] Though as Maggi (2014) observes, the preference for shallow integration over deep integration, if both achieve efficiency, is typically presumed but not modeled.

[13] As Staiger and Sykes (2013) explain, the nonviolation complaint was successfully applied to prevent domestic subsidies to import-competing industries in the early years of the GATT, well before such subsidies were explicitly ruled out in the WTO. Such successful complaints have not continued under the WTO, but Staiger and Sykes (2017) offer a theory for how such complaints still enhance efficiency even if we do not observe them regularly.

the GATT/WTO history is consistent with nations avoiding deep agreements for policies affecting import-competing industries, and there is less evidence members are avoiding deep agreements that limit domestic policies affecting export industries.

To summarize, limiting the instruments under consideration to avoid cross-country redundancy facilitates the existence of a stable agreement when reciprocal negotiations can address both terms-of-trade externalities and local price externalities in achieving efficiency and stability. If prohibiting domestic policies that benefit import-competing industries is too costly, norms and rules that limit policies that promote export industries is a viable alternative, and this is consistent with the institutional history of the GATT/WTO.

4.3 Local Price Effects at the Stable Outcome

To confirm that trade agreements address local price externalities, the signs of local price effects at stable trade agreement outcomes are a potential source of evidence. Notice that Proposition 5 implies two new types of stable outcomes under reciprocity: one in which the local price effects on government objectives are negative, and one in which they are positive.

To interpret these local price effects, first consider their sign at the noncooperative policies. From (19), we find at noncooperative policies that $\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} < 0$ (we omit similar results for Foreign), because the negative effects through local prices counter the positive effects of policy on government objectives from manipulating the terms of trade ($\frac{dW}{dp^w}\frac{dp^w}{d\Lambda} > 0$).

We have shown that nations gain from reciprocal policy reductions (Remark 1), and the inequalities (16) ensure a link between reciprocal policy reductions and trade expansion. When there is no role for trade agreements to address local price externalities, the cooperative equilibrium involves reciprocal cooperation that expands trade to the level of the Bagwell and Staiger political optimum, such that there are no local price effects on government objectives ($\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} = 0$). For stable and efficient agreements such that local price externalities matter, we can only confirm that the *total* local price effects are zero (as in equation 25). The possibility that agreements address local price externalities then offers two alternative possibilities: (1) local price effects of policy are still negative at the stable equilibrium ($\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} < 0$), and trade expansion is less than the politically optimal level, and (2) local price effects are positive ($\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} > 0$), and trade expansion is beyond the politically optimal level.

Of the two new possible outcomes, we focus less on the possibility that cooperative trade expansion at the stable outcome is short of the politically optimal level. For this case to exist, nations must be harmed (apart from terms-of-trade gains) when they export more as their trading partner reduces trade barriers. Regardless of whether negative externalities from exporting are practically relevant or not, there is nothing new in theory about finding that

$\frac{dW}{dp^l}\frac{dp^l}{d\Lambda} < 0$. This is the case for any policies on the liberalization path between noncooperative policies and the political optimum.

So we focus instead on the possibility of trade expansion beyond the politically optimal level, and we focus on specific examples of this possibility in Section 5. Other theories mentioned thus far do not explain why nations would prefer a local price increase in import-competing sectors after an agreement, and such an outcome cannot occur on a liberalization path to the Bagwell and Staiger political optimum.

The possibility that nations value local price increases at stable trade agreements outcomes is of significant importance both for empirical work and legal scholarship. There is now a wide body of empirical work studying how non-tariff measures can create additional barriers to trade (see e.g. Ederington and Ruta, 2016). Such non-tariff measures increase costs of imports while providing no revenue like import tariffs would. Why then would countries impose such barriers? Literature measuring effects of these barriers typically does not ask whether it is politically rationale to impose them. One possibility is a type of terms-of-trade manipulation such that nations could use domestic regulation excessively to shift costs of legitimate regulatory goals onto trading partners (Staiger and Sykes, 2011; Staiger, 2015). If such a possibility is not explicitly acknowledged, then the presumption must be that discriminatory protectionist barriers are somehow politically desirable if they raise prices for foreigners and favor the domestic import-competing industry. But to reiterate, such a local price increase alone cannot be desirable anywhere on the liberalization path according to the prior theory of stable trade agreement outcomes in Bagwell and Staiger (2016). What theory of stable policies in the current paper offers is the possibility that local price effects are positive at the stable outcome, and in this case, a protectionist non-tariff measure that increases a domestic local price can be desirable regardless of terms-of-trade motives.

The possibility of positive local-price effects at a stable equilibrium also addresses a long-standing criticism by legal scholars and trade practitioners of the economic theory of trade agreements. A recent illustration of this criticism comes from Regan (2015). He argues that trade agreements do not address terms-of-trade manipulation. Instead "trade agreements are about reducing politically motivated protectionism; and getting an agreement depends on political support from exporters." Prior theory does allow for welfare of importers to be traded off against welfare of exporters in negotiations when trade agreements only address terms-of-trade manipulation, so prior theory leads Grossman (2016) to conclude that Regan's criticism is more "semantics than substance." But there are multiple parts to Regan's claim that need to be considered separately and carefully.

Based on recent theory and empirics, the weakest part of Regan's claim is that trade agreements do not address terms-of-trade manipulation. A wide body of theory, including the current paper, finds that terms of trade motives yield higher tariffs at noncooperative policies. There is now a wide body of evidence that trade agreements do in fact eliminate a component of noncooperative tariffs that is motivated by terms-of-trade manipulation (Bagwell and Staiger, 2011; Ludema and Mayda, 2013; Blanchard, Bown, and Johnson, 2016). So the economics literature provides no support for this portion of Regan's criticism.

The stronger part of Regan's argument relates to the empirical claim of practitioners that trade agreements also reduce politically-motivated protectionism. He proceeds to discuss how this claim runs counter to prior theory of trade agreements addressing solely terms-of-trade externalities, and he is correct on that point. If we take politically-motivated protectionism to mean the component of tariffs that is imposed for purely redistributional purposes, then indeed the prior theory of trade agreements addressing solely terms-of-trade externalities does not offer an explanation for why such protectionist tariffs would be reduced.[14] So if practitioners are correct that trade agreements reduce the politically-motivated component of tariffs, then Regan's criticism is in fact substantial, not semantic.

The current paper addresses the substantial part of Regan's criticism by deriving a stable equilibrium in which local price externalities are positive. A transition from the noncooperative equilibrium to the political optimum (where local price effects are zero) is consistent with the elimination of policy motivated by terms-of-trade manipulation. A transition from the political optimum to the new stable equilibrium allows for the possibility that nations cooperate to address local price externalities in addition to terms-of-trade manipulation. As for Regan's claim that the political organization of exporters relates to the reduction of such "protectionist" tariffs, we come back to theory and evidence supporting this claim after introducing politically-organized exporters in the next section.

5 Applications

We proceed to apply stability under reciprocity to specific models in which local price externalities matter. The first environment that we consider is the partial equilibrium, perfectly competitive model of Bagwell and Staiger (2001b). We first confirm that the model fits the assumptions laid out in Section 3.1. As Bagwell and Staiger (2016) first show, there is a local price externality if nations have only import policies (i.e. limited instruments) and exporters are politically organized. We find that there are stable policies that yield the same local prices as the stable policies in the setting with import and export policies, and either set of stable policies is Pareto efficient.

[14] As mentioned in the introduction, Maggi and Rodriguez-Clare (2007) do model how such tariffs could be reduced if trade agreements address both terms-of-trade externalities and commitment problems.

At the stable equilibrium in the limited-instrument setting, nations prefer a positive increase in the local price. In the context of the model, the interpretation is that nations liberalize so much that the political losses for domestic import-competing industries outweigh gains for consumers from lower prices. To reiterate, this positive local price derivative does not occur on the liberalization path if trade agreements address only terms-of-trade externalities.

We then focus on the empirical implications of nations desiring an increase in the local price. Stronger political organization from exporters implies lower tariffs from cooperative agreements, and Ludema and Mayda (2013) imply empirical support for this claim. Sufficiently strong importer organization implies that rules preventing export policies would take the form of export subsidy restrictions at the stable and efficient outcome. The model rationalizes how smaller countries would be willing to impose contingent protection or disguised non-tariff protection even if they have no market power that would permit them to manipulate their terms of trade.

The remainder of the section illustrates the stable policies under reciprocity using specific functional forms. First, we consider specific functional forms for the model of Bagwell and Staiger (2001b) with only import tariffs. Second, we consider reciprocity between increasing the final good price and promoting intermediate exports in Antràs and Staiger (2012b). Lastly, we consider reciprocal increases in final good prices in a symmetric version of Antràs and Staiger (2012b) in which there are no intermediate trade policies.

5.1 The Perfect Competition Model

This subsection follows Bagwell and Staiger (2001b, 2016). We maintain consistent notation except for some minor modification to ensure consistency with the rest of the current paper. There is an economy with goods x and y, such that Home imports x and exports y, and there is a freely traded outside numeraire good that enters into welfare quasilinearly. The political economy objectives for Home and Foreign (* superscript) are

$$
\begin{aligned}
W(p_x, p_y^*, p_x^w, p_y^w) &= \int_{p_x}^{\bar{p}} D(p_x^1)dp_x^1 + \gamma_M \Pi_x(p_x) + (p_x - p_x^w)M_x(p_x) \\
&+ \int_{p_y(p_y^*)}^{\bar{p}} D(p_y^1)dp_y^1 + \gamma_E \Pi_y(p_y(p_y^*)) - (p_y(p_y^*) - p_y^w)M_y(p_y^*), \text{ and} \\
W^*(p_x, p_y^*, p_x^w, p_y^w) &= \int_{p_x^*(p_x)}^{\bar{p}} D(p_x^1)dp_x^1 + \gamma_E^* \Pi_x^*(p_x^*(p_x)) - (p_x^*(p_x) - p_x^w)M_x(p_x) \\
&+ \int_{p_y^*}^{\bar{p}} D(p_y^1)dp_y^1 + \gamma_M^* \Pi_y^*(p_y^*) + (p_y^* - p_y^w)M_y(p_y^*),
\end{aligned} \tag{27}
$$

such that D is demand (a decreasing function), Π_x and Π_y are profits, and M_x and M_y are import demand functions. The objective includes standard political economy weights γ_M, γ_E, γ_M^*, and γ_E^* which are all greater than one. In our earlier vector notation, $M = \{-M_x, M_y, Z\}$ where Z is Home imports of the outside good, $p^l = p^{l*} = \{p_x, p_y^*, 1\}$, and $p^w = \{p_x^w, p_y^w, 1\}$, and the balanced trade condition is still $Mp^w = 0$.

For policy, Home chooses import tariff τ_x and Foreign chooses import tariff τ_y^*, and these tariffs are chosen to be nonprohibitive. We exclude export policies. There are increasing supply functions $Q_x(p) = Q_y^*(p) < Q_y(p) = Q_x^*(p)$. Under profit maximization, $\frac{d\Pi_x}{dp_x} = Q_x(p)$, and similar derivatives hold for the other profit functions. To close the model we require the no-arbitrage conditions and market clearing conditions such that

$$
\begin{aligned}
p_x - \tau_x &= p_x^* = p_x^w, \\
p_y^* - \tau_y^* &= p_y = p_y^w, \text{ and} \\
Q_i(p_i) + Q_i^*(p_i^*) &= D(p_i) + D(p_i^*) \text{ for } i = x, y.
\end{aligned} \tag{28}
$$

Notice that under these conditions, specifying either of the prices or the tariff (e.g. p_x, p_x^*, p_x^w, or τ_x) fully determines the other variables for that good. Thus we can define W as a function of $p_y(p_y^*)$ instead of p_y, and W^* as a function of $p_x^*(p_x)$ instead of p_x^*.

We verify that the model still fits our assumptions of Section 3.1, so we can later apply our general results. The equations (28) ensure that higher import tariffs imply higher local prices in the import market and lower prices in the export and world markets, thus ruling out Metzler and Lerner paradoxes as required. And M_x and M_y are both decreasing in the local price of imports (since a higher price decreases demand D_i and increases supply Q_i for $i = x, y$), so (15) and (16) hold. Then from (27), the import tariffs must improve the terms of trade for the nation imposing them, so (14) holds. The terms of trade effects satisfy (17), as $\frac{dW}{dT} = 1$ and $\frac{dW^*}{dT^*} = -1$ in the quasilinear setting. Without the export policies, there is a local price externality, because Home lacks an instrument to affect p_y^* and Foreign lacks an instrument to affect p_x. We can verify

$$\frac{dW}{dp_y^*} = \frac{dp_y}{dp_y^*}(\gamma_E - 1)Q < 0, \tag{29}$$

because $\gamma_E > 1$ and

$$\frac{dp_y}{dp_y^*} = \frac{M_y'}{Q_y' - D_y'} = \frac{p_y^* \mu_y}{p_y \xi_y} < 0,$$

for import demand elasticity μ_y and export supply elasticity ξ_y. So a Foreign tariff decrease allows Home to benefit not only via a terms-of-trade gain, but also via a lower p_y^* and higher p_y, since there are higher weighted profits for Home's exporters of y. Home benefits even though there are higher consumer prices. Notice that this externality could be defined either as local price externality abroad, or (as in Bagwell and Staiger, 2016) as a domestic local price externality, because (28) implies one price pins down the other. The local price externality here amplifies the terms-of-trade externality, so $\frac{dW}{d\tau_y} < 0$. Similarly for Foreign, $\frac{dW^*}{dp_x} < 0$ and $\frac{dW}{d\tau_x} < 0$. Thus, this model fits into the general framework of Section 3.1.

Now we can interpret the model through the results of our previous sections. In this setting, there are only two independent local prices to target and two policies to be determined by the agreement. Thus, any more instruments such as export policies do not provide any more gains in efficiency, by the argument in Proposition 7. If nations have both import and export policies, then the conditions for Proposition 4 are satisfied, so the unique stable and efficient policy is the political optimum. The efficiency of the political optimum here is consistent with what Bagwell and Staiger (2001b, 2016) have already shown.

For reciprocity satisfying the standard rule as in (23), we have

$$\frac{d\tau_y^*}{d\tau_x} = \frac{\frac{dM_x}{dp_x}\frac{dp_x}{d\tau_x}p_x^w}{\frac{dM_y}{dp_y^*}\frac{dp_y^*}{d\tau_y^*}p_y^w} = \frac{M_x \frac{dp_x^w}{d\tau_x}}{M_y \frac{dp_x^w}{d\tau_y^*}} > 0 \tag{30}$$

for reciprocal tariff decreases. The stable efficient point then consists of the policies $(\hat{\tau}_x, \hat{\tau}_y^*)$ satisfying the following

$$\begin{aligned} \frac{dW}{dp_x}\frac{dp_x}{d\tau_x} + \frac{dW}{dp_y^*}\frac{dp_y^*}{d\tau_y^*}\frac{d\tau_y^*}{d\tau_x} &= 0, \text{ and} \\ \frac{dW^*}{dp_y^*}\frac{dp_y^*}{d\tau_y^*}\frac{d\tau_y^*}{d\tau_x} + \frac{dW^*}{dp_x}\frac{dp_x}{d\tau_y^*} &= 0. \end{aligned} \tag{31}$$

By Proposition 1, these policies are efficient, and by Proposition 7, they obtain the same efficiency frontier as the model with both import and export policies. From (15), (29), (30), and (31), we have $\frac{dW}{dp_x} > 0$ and $\frac{dW^*}{dp_y^*} > 0$. We summarize the results as follows.

Remark 2 *For the perfect-competition, partial-equilibrium trade model, there are efficient policies (defined over the space of both import and export policies) such that the policies are stable with respect to reciprocal negotiations over only import policies. At these policies, both nations desire an increase in the price in their import-competing sector.*

For stability under reciprocity to be achieved, the net political gains for the export industry must be offset by domestic losses, such that Home and Foreign are each losing domestically by cutting their tariffs further. Notice then that

$$\frac{dW}{dp_x} = (\gamma_M - 1)Q + \tau_x M'_x. \tag{32}$$

Meanwhile, the Nash tariff is such that

$$\frac{dW}{dp_x}\frac{dp_x}{d\tau_x} = \frac{dp_x^w}{d\tau_x}M, \tag{33}$$

and since $\frac{dp_x}{d\tau_x} > 0 > \frac{dp_x^w}{d\tau_x}$, the tariff is set sufficiently high so that $\frac{dW}{dp_x} < 0$, because Home pursues terms-of-trade gains. If nations had import and export policies, we would instead have $\frac{dW}{dp_x} = 0$ so $\tau_x = \frac{(\gamma_M - 1)Q}{-M'_x} > 0$ for $\gamma_M > 1$, and this is the usual Bagwell and Staiger political optimum. In the current model with only import tariffs, we instead have the tariff τ_x set so low that $\frac{dW}{dp_x} > 0$, and τ_x is less than the politically optimal level.

To interpret the stable equilibrium where $\frac{dW}{dp_x} > 0$, notice that the negotiation is balancing the political economy gains for the import-competing industry, the export industry, and the consumers. At the equilibrium, the political-weighted losses to the import-competing industry from lower prices and the loss in tariff revenue must be outweighing the gains to consumers from lower prices.

5.2 Evidence

We have previously derived the new possibility of an equilibrium such that each nation desires a local price increase in the import-competing sector. Having introduced a specific model, we can now discuss several empirical implications for this new possibility and what evidence there is to support them.

The first point of evidence relates to the role of exporter political economy in tariffs. From (29), (31), (32), (33). we can conclude the following:

Remark 3 *For the perfect-competition, partial-equilibrium trade model, exporter political economy γ_E/ξ_y (political organization over export supply elasticity) affects both the stable tariffs under reciprocity and the size of the decrease in tariffs from the Nash level, but does not affect the level of Nash tariffs.*

Notice the contrast between this result and the effects of importer political organization, which implies larger tariffs at both the Nash equilibrium and the stable outcome. This result also stands in contrast to the full instrument model (as in Bagwell and Staiger, 2001b) in which case exporter political economy affects both the Nash and cooperative levels, but not the decrease. As for evidence, Ludema and Mayda (2013, Table 3) include exporter organization times inverse export supply elasticity as a control when estimating terms-of-trade effects in a similar model, and they do find that larger exporter organization implies lower import tariffs following the Uruguay Round of trade negotiations.[15] Their result is consistent with trade agreements addressing local price externalities and not only terms-of-trade externalities.[16] The new theory of the current paper is what enables this interpretation.

The second point of evidence relates to the WTO prohibition of export subsidies. As Sykes (2005) and DeRemer (2013a) detail, the GATT/WTO in practice started to limit export subsidies for a subset of members in 1962 and then all members in 1995, minus a few phased-out exceptions for developing countries. What do the full instrument and limited instrument model imply about this? In the full instrument version, the export subsidy at the political optimum implies a terms-of-trade loss, so nations would ban export taxes rather than export subsidies, and this is a fairly typical result of this literature.[17] In the limited instrument version we can conclude the following:

Remark 4 *For the perfect-competition, partial-equilibrium trade model, there is a sufficiently large import parameter γ_M or sufficiently large export parameter γ_E such that nations prefer limits on export subsidies at the stable outcome with only import tariffs.*

To confirm this, notice that the externalities of a Foreign export subsidy s_x^* would be

$$\frac{dW}{ds_x^*} = \frac{dW}{dp_x}\frac{dp_x}{ds_x^*} - \frac{dp_x^w}{ds_x^*}M.$$

[15] Ludema and Mayda (2013) focus though on a trade agreement outcome based on the principal supplier rule and Nash bargaining, rather the current paper's trade agreement outcome based on reciprocity. The result is consistent with the trade agreement addressing local price externalities in either case.

[16] Though exporter political economy does enter into cooperative tariffs in the full instrument model of Bagwell and Staiger (2001b), this theory cannot explain the empirical result concerning exporter organization. Such an explanation would require nations to utilize export subsidies widely, but such subsidies were prohibited for most WTO members in the 1990s.

[17] See Lee (2016) for a survey. Exceptions though include Bagwell and Staiger (2012a) and Bagwell and Lee (2015) in which case there is a Metzler paradox, and export subsidies can be terms-of-trade improving.

The second term is always positive, reflecting the terms-of-trade gains from the subsidy, while the first term is negative. $\frac{dW}{dp_x} > 0$, and this reflects the negative effects of more import competition from a subsidy increase from the stable point. The only term that depends on γ_M is $\frac{dW}{dp_x}$, so by the envelope theorem (representing the equilibrium as the constrained optimization solution in Section 2.3), we have $\frac{d^2W}{ds_x^* d\gamma_M} < 0$, so for sufficiently large γ_M there is a negative externality from the subsidy. Because the stable point is efficient, $\frac{dW}{ds_x^*} < 0$ implies $\frac{dW^*}{ds_x^*} > 0$.

Similarly, looking at Foreign's policy incentives,

$$\frac{dW^*}{ds_x^*} = \frac{dW^*}{dp_x}\frac{dp_x}{ds_x^*} + \frac{dp_x^w}{ds_x^*}M$$

we can derive that $\frac{d^2W^*}{ds_x^* d\gamma_E} > 0$, and a sufficiently large γ_E ensures that Foreign would be willing to impose the export subsidy, and $\frac{dW^*}{ds_x^*} > 0$ implies $\frac{dW}{ds_x^*} < 0$ at stable and efficient policies.

To summarize, the possibility that trade agreements address local price externalities allows the possibility for a stable outcome under reciprocity such that restrictions on export policies take the form of a ban. The ban is desirable only because at the stable point, nations have liberalized so much that the loss from a local price decrease to import-competing sectors outweighs the benefits to consumers. This possibility then widens the set of models such that export subsidy restrictions can be rationalized. The export subsidy restrictions can be rationalized at stable outcomes in the limited instrument setting, while the political optimum in the full instrument setting provides no rationale for export subsidy restrictions.[18]

The last example we consider is the possibility of imposing non-tariff measures and contingent protection. Theory (Bagwell and Staiger, 1990) and empirics (Bown and Crowley, 2013) of contingent protection are based on the possibility that trade demand shocks (a shifting of some parameter in the export supply function) can lead to temporary protection if the terms-of-trade gains outweigh the local price distortion from raising the tariff. Similarly, temporary political shocks can also motivate temporary trade barriers as they would imply a higher politically optimal tariff during the period of political pressure (Bagwell and Staiger, 2005b). In the current paper, when nations have an agreement such that $\frac{dW}{dp_x} > 0$, then the contingent protection following an import shock would be rational regardless of the scope for terms-of-trade manipulation. The possibility can then more easily explain the use of contingent protection by emerging markets with limited market power, even in the absence of changes in political preferences. And as discussed in Section 4.3, an equilibrium such

[18] The results here are then similar to those in DeRemer (2013a), who uses a model with profit-shifting and monopolistic competition. But that paper considers whether subsidy restrictions are desirable when nations have agreed to zero import tariffs, rather than at a stable equilibrium under reciprocity.

that $\frac{dW}{dp_x} > 0$ implies that nations would attempt to impose non-tariff measures as disguised protection to increase prices, even if such measures neither improve terms of trade nor shift regulatory costs abroad.

5.3 Perfect Competition with Specific Functional Forms

To illustrate further, we consider specific functional forms. Following Bagwell and Staiger (2001b), $CS(p) = .5(1-p)^2$, the profit functions for the export goods are $p^2/2$, and the profit functions for the import-competing sectors are $p^2/4$. Balanced trade implies world prices are $p_x^w(\tau_x) = (4-3\tau_x)/7$ and $p_y^w(\tau_y^*) = (4-3\tau_y^*)/7$.

We can then solve for the equilibrium as a function of the political economy parameters. First we consider the case in which the political parameters are symmetric, so there is one parameter for export industries γ_E and one parameter for import-competing industries γ_M. In this case, the stable point is in fact the same as the symmetric efficient point in the limited-instrument setting. We can derive that the efficient import tariffs are

$$\tau_x = \tau_y^* = \frac{4(2\gamma_M + 1 - 3\gamma_E)}{59 - 9\gamma_E - 8\gamma_M}. \tag{34}$$

As anticipated from Proposition 7, the level of total trade barriers and local prices are the same as in the political optimum in Bagwell and Staiger (2001b) when both import policies and export policies are available, so the same level of welfare is obtained even without the export policies. So here banning export policies has no inefficiency consequences.

We depict the Nash and stable efficient equilibria graphically in Figure 1, for the case of $\gamma_M = 1.2$ and $\gamma_E = 1.1$, which implies small positive tariffs at the stable and efficient point. The curves here reflect iso-gains for the first-order welfare effects from differential reciprocal policy changes. The curves S and S^* indicate where in the policy space Home and Foreign are indifferent to reciprocal policy changes, as in equations (31). At the curves N and N^*, Home and Foreign get the same welfare from reciprocal policy changes as they do at the Nash policies. The liberalization path then involves the progression between these iso-gain curves until both Home and Foreign gain zero welfare from reciprocal policy changes. The curve EF, between S and S^*, is the Pareto efficiency frontier, and the stable outcome under reciprocity lies at the intersection of the three.

The political optimally policies under limited instruments, at which point $\frac{dW}{dp_x} = 0$ and $\frac{dW^*}{dp_y^*} = 0$, is labelled as PO. It lies in between the Nash equilibrium and the efficient stable point, because the local price externalities from politically-organized exporters imply deeper liberalization is necessary to achieve the efficiency frontier.

A more interesting case is when political organization is asymmetric. Even without the

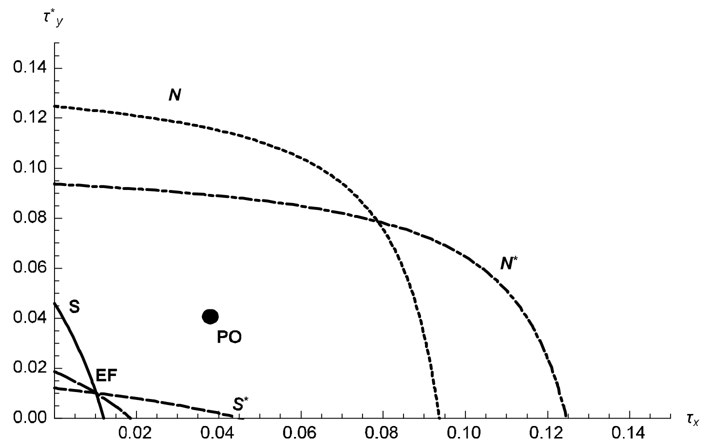

Figure 1: Symmetric Model

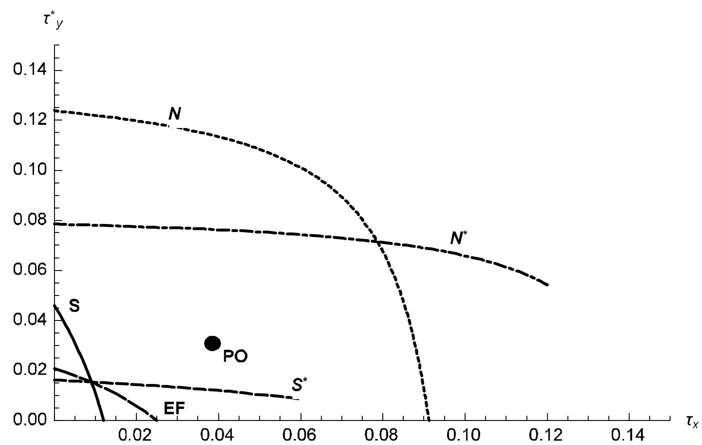

Figure 2: Asymmetric Model

theory of this paper, one could have predicted the outcome of reciprocal negotiations would be the symmetric policies on the efficiency frontier. The value of the stable point is that it yields a prediction for the outcome of negotiations even in the asymmetric case.

The plot in Figure 2 is for parameters $\gamma_M = 1.2$, $\gamma^*_M = 1.15$, $\gamma_E = 1.1$, $\gamma^*_E = 1.05$. With these parameters, the efficient point is no longer such that we maximize the global welfare objective with $\lambda = 1$. Instead the stable point maximizes an objective with $\lambda = 1.18$ weight on Foreign. Even though the political economy forces are stronger in Home, the reciprocity concept here is neutral with respect to political economy forces, and the Home tariff is smaller than the Foreign tariff at the stable efficient point. In either case, the political optimum lies in between the stable efficient point and the Nash equilibrium, and each nation benefits from a local price increase at the stable point.

5.4 The Offshoring Model

The offshoring model of Antràs and Staiger (2012b) considers Foreign, who exports intermediates, and Home, who is the final good producer. Home has a policy τ_1 (either trade or domestic) that determines the final good price of $p_1 = (1+\tau_1)$ and Foreign has no direct way to influence that price. As discussed in Section 3.4, there is an inefficient political optimum and a local price externality from changes in the final good price, provided that prices are determined by bilateral bargaining and the Foreign government places a political economic weight on the profits of the intermediate exporter. Both nations could have trade policies that influence trade in the intermediate, and they attempt to manipulate their terms of trade through all policies. Neither nation can influence the terms of trade of the final good.

As we discussed in Section 4.1, with all three policies, a stable outcome under the standard reciprocity rule does not exist. Foreign would like to address the local price externality in p_1^H through reciprocal negotiations involving intermediate good policy, but this cannot be done if both nations can influence the local prices of intermediate, by Proposition 6. The alternative is to have Home liberalize trade in the intermediate fully, and to allow reciprocal negotiations involving Foreign's policy promoting exports of the intermediate and Home's policy affecting the final good price. Reciprocal negotiations can then allow each nation to internalize the profits over the whole supply chain. This limitation in policies in negotiations comes at no loss to efficiency by Proposition 7.

The details of the model follow Antràs and Staiger, though we now take the Home tariff on the imported intermediate to be zero. There is a production function $y(x)$ for the final good, and the price of the intermediate is determined by bilateral bargaining. The Home producer has bargaining power $\alpha \in (0,1)$ in the Nash bargaining game. The equilibrium intermediate trade volume $x(p_1, p_H^x, p_F^x, p_w^x)$ is determined by Foreign intermediate producers equating their marginal benefit to their marginal cost normalized to 1:

$$(1-\alpha)(p_1 y'(x) - p_H^x + p_F^x) = 1.$$

The prices for the intermediate x, when Foreign has export policy τ_F^x, are as follows:

$$\begin{aligned} p_w^x &= (1-\alpha)\frac{y p_1}{x} + \alpha \tau_F^x, \text{ and} \\ p_F^x &= p_w^x - \tau_F^x, \text{ and } p_H^x = p_w^x. \end{aligned}$$

The Home and Foreign objectives are

$$\begin{aligned} W_H &= \int_{p_1}^{\bar{p}} D(\tilde{p}_1)d\tilde{p}_1 + \gamma_H(yp_1 - xp_H^x) + (p_1 - 1)(D(p_1) - y) + (p_H^x - p_w^x)x, \text{ and} \\ W_F &= \gamma_F(p_F^x - 1)x + (p_w^x - p_F^x)x, \end{aligned}$$

where D is the Home demand function for the final good, and γ_H and γ_F are political economic weights on profits in each country. The final good is consumed only in Home.

The objectives in terms of prices are $W(p_1, p_H^x, p_F^x, p_w^x)$ and $W^*(p_1, p_H^x, p_F^x, p_w^x)$. Home, when using policy to set p_1, will not internalize the benefit of the higher final good price for Foreign, while Foreign will manipulate its terms of trade with the export policy. The reciprocal cooperation undoes both terms of trade externalities and local price externalities. The stable outcome under reciprocity will be efficient according to Proposition 1.

Results in this model hinge on the form of the production function. Here we use the function $y = 5\log(1 + x)$, where recall y is final good production and x is intermediate production. As Antràs and Staiger (2012b) observe for this functional form, the higher final good tariff worsens the terms of trade of Home in importing the intermediate, whose price rises.[19] We can then consider reciprocity that involves Home increasing the price of the final good (while improving Foreign's terms of trade for the intermediate) and Foreign progressing from an export tax to an export subsidy that leads to more efficient trade in the intermediate (while worsening Foreign's terms of trade for the intermediate).

We illustrate model outcomes for specific parameters and functional forms. We assume linear demand of $2-p$ and bargaining power equally split between each nation's firm ($\alpha = .5$). Home maximizes national income while Foreign has a political economy weight $\gamma_F = 1.1$. As mentioned at the outset, one of the values of the framework is in selecting among points on the efficiency frontier in asymmetric environments. We find in this case that the stable equilibrium under reciprocity selects an outcome that corresponds to a weight λ on Foreign objectives that is two percent higher than the weight on Home objectives. Table 1 shows the policies, prices, and trade volume for the Nash equilibrium under noncooperative policies and the efficient stable point under reciprocity.

[19] Another possibility is that the higher final good price could decrease intermediate trade volume and worsen terms of trade for the intermediate supplier. In this case, the nation of the intermediate supplier would instead reciprocally negotiate for a lower final good price.

Table 1		
	Nash point	Stable point
τ_H^1	.03	.30
τ_F^x	.58	-.74
p_w^x	2.08	.91
p_F^x	1.50	1.65
x	1.00	4.17

At the stable efficient point relative to the Nash equilibrium, Home chooses a higher final good price, and Foreign promotes rather than taxes its exports. Reciprocal negotiations achieve efficiency by increasing trade volume in the intermediate.

5.5 The Global Supply Chain Model

Blanchard, Bown, and Johnson (2016) detail a model of global supply chains in which trade in intermediates and value-added factors matter, but the only trade instruments available are import tariffs on final goods. In this setting, as in Antràs and Staiger (2012b), there are local price externalities. When policymakers set prices of final goods, foreign-value added (or foreign supply of intermediates) matters in the price-setting decision. So gains from protection of the final good go to the foreigners, as do losses from liberalizing the final good. This global externality does not show up in either the optimal tariff formula or the formula for the political optimum. The empirical implementation in Blanchard, Bown, and Johnson (2016) uses the prediction that the political optimum is the outcome of bilateral negotiations, meaning that trade agreements address only terms-of-trade externalities. As we have discussed in Section 3.4, in the presence of these local price externalities the political optimum is inefficient. Though the political optimum is still a plausible outcome of bilateral negotiations if local price externalities are minimal, the political optimum is not a stable outcome of bilateral negotiations under reciprocity. Our theory here provides an alternative.

We illustrate using a symmetric extension of the Antràs and Staiger model in which nations engage in reciprocal reductions of the final good tariff and have no trade policies in intermediates. Each nation has one final good producer and one intermediate supplier for the other's final good producer, following the functional forms from the offshoring model in the previous subsection. Both nations will set the final good prices too low because they do not internalize the benefits of the final good price that flow upstream. The production function is still $y = 5\log(1+x)$, in which case higher final good tariffs improve the terms of trade of the intermediate exporter. Table 2 details outcomes of Nash policies, politically optimal policies (for this setting limited to final good import tariffs), and stable policies.

Table 2			
	Nash policies	Political optimum	Stable policies
W	4.73	5.52	5.53
W_1	2.79	1.39	1.18
W_x	1.93	4.13	4.35
τ^1	.025	0.68	0.74
p_w^x	1.54	1.89	1.90
x	1.56	3.22	3.35

In Table 2, W_1 is the welfare from each nation's final good sector and W_x is the welfare from each nation's intermediate sector. Cooperation involves higher final good prices (higher τ^1) and shifting welfare from the final good sector (W_1) to intermediate production (W_x) while expanding trade volume. With the chosen parameters, the political optimum is only marginally inefficient, but relative to the stable point under reciprocity, the political optimum underestimates the final good tariff and the trade volume of the intermediate. The predicted larger expansion of trade at the stable efficient point relative to the political optimum is consistent with the previous examples in which local price externalities matter for trade cooperation.

Interestingly, cooperation takes the form of increasing external trade barriers, even if there is no ability to manipulate external world prices. Reciprocity increases internal trade liberalization and decreases external liberalization, though this result hinges on having assumed a production function and bargaining process such that higher final good prices imply higher profits upstream.

6 Conclusion

This paper contributes to the theory of trade agreements by identifying the general efficiency properties of a particular stable policy set, such that neither nation would deviate using reciprocal policy changes. These stable policies provide a predictive outcome for trade negotiations when local price externalities matter for global efficiency. While other work considers reciprocity rules and cooperation in trade negotiations when local price externalities matter, this work has typically done so within the context of specific market structures, while this paper considers more broadly how reciprocity rules could function. We then apply the general theory to particular models, and we obtain insightful predictions for trade cooperation in both the standard partial equilibrium model and in the offshoring setting. We next discuss directions for future research.

The standard theory of reciprocity focuses on understanding the starting point and end points of reciprocal liberalization. In more asymmetric environments, the transition between the two can be problematic if, for example, reciprocity falls along an iso-world-price line and the starting point and end point have different world prices. There is need for better understanding about the transition between world prices in these situations.[20]

Another frontier involves alternative reciprocity rules. In a symmetric environment, the reciprocity rule is immaterial if reciprocity is also symmetric and simply proceeds until the end point. In asymmetric environments, one nation could cease to prefer reciprocal policy changes before reaching the efficiency frontier. Thus, reciprocity rules could then differ in terms of feasible starting points for achieving the efficiency frontier via reciprocal policy changes. The alternative rules in the literature by Mrázová (2011) and Ossa (2011) preserve trade balance within sectors. Such a rule could narrow the set of permissible reciprocal policy changes. This could be limiting, but could also be desirable if it rules out policy changes that are not mutually beneficial. A worthwhile effort would be to formulate whether the potential advantage of such a rule outweighs the disadvantage.

Lastly, we discuss the paper's lessons for the importance of classifying international externalities. Grossman (2016), for example, argues that the literature's effort to "pin labels" on international externalities is of unclear importance, so effort spent instead on questions of trade agreement design would be more productive. The exercises in this paper illustrate the value from classifying externalities appropriately, because this classification matters for trade agreement design. What we have shown is that local price externalities can lead to new design problems and new predictions for trade negotiation outcomes.

A Appendix

Proposition 6 *Suppose that for any policies, there is some $p_i \in p^l \cap p^{l*}$, such that each nation can influence p_i independently from other prices, and p_i is not fully determined by p^w. Then if $(\Lambda^{R^0}, \Lambda^{*R^0})$ exists, it is equivalent to $(\Lambda^{PO}, \Lambda^{*PO})$. If $(\Lambda^{PO}, \Lambda^{*PO}) \notin EF$, then no stable point under the standard reciprocity rule exists.*

Proof. Suppose we are at $(\Lambda^{R^0}, \Lambda^{*R^0})$. Re-express $\frac{dp^l}{d\Lambda}$ and $\frac{dp^l}{d\Lambda}$ over a different basis $\dot{\Lambda}$ and $\dot{\Lambda}^*$ such that $\frac{dp^l}{d\dot{\Lambda}_1} = \frac{dp_i}{d\dot{\Lambda}_1} \neq 0$ and $\frac{dp^{l}*}{d\dot{\Lambda}_1^*} = \frac{dp_i}{d\dot{\Lambda}_1^*} \neq 0$. Then for $\frac{d\dot{\Lambda}_1^*}{d\dot{\Lambda}_1}$, which is consistent with changes in $\dot{\Lambda}$ and $\dot{\Lambda}^*$ satisfying R^0, we have $\frac{dW}{d\dot{\Lambda}_1} + \frac{d\dot{\Lambda}_1^*}{d\dot{\Lambda}_1}\frac{dW}{d\dot{\Lambda}_1^*} = \frac{dW}{dp_i}\left(\frac{dp_i}{d\dot{\Lambda}_1} + \frac{d\dot{\Lambda}_1^*}{d\dot{\Lambda}_1}\frac{dp_i}{d\dot{\Lambda}_1^*}\right)$. The independence from world price implies the R^0-preserving change in policies does not also preserve p_i. For home not to gain from any reciprocal changes, the only possibility is $\frac{dW}{dp_i} = 0$. This and our

[20] Bagwell and Staiger (2018) explain this shift to the politically-optimal world prices through a theory of dominant strategies in bargaining.

construction of $\dot{\Lambda}_1$ implies $\frac{dW}{dp^l}\frac{dp^l}{d\dot{\Lambda}_1} = \frac{dW}{dp_i}\frac{dp^i}{d\dot{\Lambda}_1} = 0$. By the same argument of Proposition 5, $\frac{dW}{dp^l}\frac{dp^l}{d\dot{\Lambda}} = 0$ and $\frac{dW}{dp^l}\frac{dp^l}{d\dot{\Lambda}^*} = 0$ for all elements of the basis we have constructed. These statements can only hold if $\frac{dW}{dp^l} = 0$, because $\dot{\Lambda}$ and $\dot{\Lambda}^*$ fully determine p^l. A similar argument implies $\frac{dW^*}{dp^{l*}} = 0$. The policies then satisfy the conditions for the Bagwell and Staiger political optimum. If $(\Lambda^{PO}, \Lambda^{*PO}) \notin EF$, then Proposition 1 would imply a contradiction, and the $(\Lambda^{R^0}, \Lambda^{*R^0})$ we supposed cannot exist. ■

References

Antràs, P. and R. W. Staiger (2012a), "Trade Agreements and the Nature of Price Determination," *American Economic Review Papers and Proceedings* 102, 470-76.

Antràs, P. and R. W. Staiger (2012b), "Offshoring and the Role of Trade Agreements," *American Economic Review* 102, 3140-83.

Bagwell, K., C. P. Bown, and R. W. Staiger (2016), "Is the WTO Passé?" *Journal of Economic Literature* 54, 1125-1231.

Bagwell, K. and Lee, S. H. (2015), "Trade Policy under Monopolistic Competition with Firm Selection," Stanford University, mimeo., April.

Bagwell, K. and R. W. Staiger (1990), "A Theory of Managed Trade," *American Economic Review* 80, 779–95.

Bagwell, K. and R. W. Staiger (1999), "An Economic Theory of GATT," *American Economic Review* 89, 215-48.

Bagwell, K. and R. W. Staiger (2001a), "Domestic Policies, National Sovereignty, and International Economic Institutions," *Quarterly Journal of Economics* 116, 519-62.

Bagwell, K. and R. W. Staiger (2001b), "Reciprocity, non-discrimination and preferential agreements in the multilateral trading system," *European Journal of Political Economy* 17, 281-325.

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

Bagwell, K. and R. W. Staiger (2005b), "Enforcement, political pressure, and the general agreement on tariffs /world trade organization escape clause," *Journal of Legal Studies* 34, 471–513.

Bagwell, K. and R. W. Staiger (2011), "What Do Trade Negotiators Negotiate About? Empirical Evidence from the World Trade Organization," *American Economic Review* 101, 1238-73.

Bagwell, K. and R. W. Staiger (2012a), "The Economics of Trade Agreements in the Linear Cournot Delocation Model," *Journal of International Economics*, 88, 32-46.

Bagwell, K. and R. W. Staiger (2012b), "Profit Shifting and Trade Agreements in Imperfectly Competitive Markets," *International Economic Review* 53, 1067-1104.

Bagwell, K. and R. W. Staiger (2015), "Delocation and Trade Agreements in Imperfectly Competitive Markets," *Research in Economics* 69, 132-156.

Bagwell, K. and R. W. Staiger (2016), "The Design of Trade Agreements," in *The Handbook of Commercial Policy*, K. Bagwell and R. W. Staiger (eds.), Elsevier, http://dx.doi.org/10.1016/bs.hescop.2016.04.005.

Bagwell, K. and R. W. Staiger (2018), "Multilateral Trade Bargaining and Dominant Strategies," *International Economic Review* 59, 1785-1824.

Baldwin, R. (2013), "Trade and Industrialization after Globalization's Second Unbundling," in *Globalization in an Age of Crisis: Multilateral Economic Cooperation in the Twenty-First Century,* R.C. Feenstra and A.M. Taylor (eds.), University of Chicago Press, 165-212.

Baldwin, R. (2016), "The World Trade Organization and the Future of Multilateralism," *Journal of Economic Perspectives* 30, 95-116.

Bergin, P. and G. Corsetti (2013), "International Competitiveness and Monetary Policy: Strategic Policy and Coordination with a Production Relocation Externality," NBER Working Paper No. 19356.

Bhagwati, J. and V. K. Ramaswami (1963), "Domestic Distortions, Tariffs and the Theory of Optimum Subsidy," *Journal of Political Economy* 71, 44-50.

Blanchard, E. J., C. P. Bown, and R. C. Johnson (2016), "Global Supply Chains and Trade Policy," NBER Working Paper No. 21883.

Blanchard, E. J. (2010), "Reevaluating the Role of Trade Agreements: Does Investment Globalization Make the WTO Obsolete?" *Journal of International Economics* 82, 63-72.

Bown, C. P., and M. A. Crowley (2013), "Self-Enforcing Trade Agreements: Evidence from Time-Varying Trade Policy," *American Economic Review* 103, 1071-90.

Campolmi, A., H. Fadinger, and C. Forlati (2014), "Trade Policy: Home Market Effect versus Terms-of-Trade Externality," *Journal of International Economics* 93, 92-107.

Cole, M. T., J. Lake, and B. Zissimos (2018), "Contesting an International Trade Agreement," mimeo., January.

Costinot, A., A. Rodriguez-Clare, and I. Werning (2016), "Micro to Macro: Optimal Trade Policy with Firm Heterogeneity," NBER Working Paper No. 21989.

Crowley, M. A. (2019), *Trade War: The Clash of Economic Systems*, CEPR Press.

DeRemer, D. R. (2013a), "The Evolution of International Subsidy Rules," ECARES Working Paper 2013-45.

DeRemer, D. R. (2013b), "Domestic Policy Coordination in Imperfectly Competitive Markets," ECARES Working Paper 2013-46.

DeRemer, D. R. (forthcoming), "Opportunities for Cooperation in Removing Prohibitive Trade Barriers," in *The World Trade Organization and Economic Development*, Ben Zissimos (ed.), Cambridge: MIT Press.

Dixit, A. (1985), "Tax Policy in Open Economies," in *Handbook of Public Economics*, A. J. Auerbach and M. Feldstein (eds.), 313-374.

Ederington, J. and J. Minier (2008), "Reconsidering the empirical evidence on the Grossman-Helpman model of endogenous protection," *Canadian Journal of Economics* 41, 501–516.

Ederington, J. and M. Ruta (2016), "Nontariff Measures and the World Trading System," in *The Handbook of Commercial Policy*, K. Bagwell and R. W. Staiger (eds.), Elsevier, http://dx.doi.org/10.1016/bs.hescop.2016.04.010

Epifani, P. and G. Gancia (2011), "Trade, markup heterogeneity and misallocations," *Journal of International Economics* 83, 1–13.

Grossman, G. M. (2016), "The Purpose of Trade Agreements," in *The Handbook of Commercial Policy*, K. Bagwell and R. W. Staiger (eds.), Elsevier, http://dx.doi.org/10.1016/bs.hescop.2016.04.016

Horn, H., G. Maggi, and R.W. Staiger (2010), "Trade Agreements as Endogenously Incomplete Contracts," *American Economic Review* 100, 394-419.

Irwin, D. A. (2004), "The Aftermath of Hamilton's 'Report on Manufacture," *The Journal of Economic History* 64, 800-821.

Johnson, H.G. (1953-54), "Optimum Tariffs and Retaliation," *Review of Economic Studies* 21, 142-153.

Lee, G. (2016), "Subsidies and Countervailing Duties," in *The Handbook of Commercial Policy*, K. Bagwell and R. W. Staiger (eds.), Elsevier, http://dx.doi.org/10.1016/bs.hescop.2016.04.009

Limão, N. (2016), "Preferential Trade Agreements," in *The Handbook of Commercial Policy*, K. Bagwell and R. W. Staiger (eds.), Elsevier, http://dx.doi.org/10.1016/bs.hescop.2016.04.013

Ludema, R. D. and A. M. Mayda (2013), "Do Terms- of-Trade Effects Matter for Trade Agreements? Evidence from WTO Countries," *Quarterly Journal of Economics* 128, 1837-93.

Maggi, G. (2014), "International Trade Agreements," in G. Gopinath, E. Helpman and K. Rogoff (eds.), *The Handbook of International Economics* 4, Elsevier, 317-90.

Maggi, G. and A. Rodriguez-Clare (2007) , "A Political Economy Theory of Trade Agreements," *American Economic Review* 97, 1374-1406.

McLaren, J. (2016), "The Political Economy of Commercial Policy," in *The Handbook of Commercial Policy*, K. Bagwell and R. W. Staiger (eds.), Elsevier, http://dx.doi.org/10.1016/bs.hescop.2016.04.001

Mrázová, M. (2011), "Trade agreements when profits matter," mimeo., February.

Ossa, R. (2011), "A New Trade Theory of GATT/WTO Negotiations," *Journal of Political Economy* 199, 122-152.

Ossa, R. (2012), "Profits in the "new trade" approach to trade negotiations," *American Economic Review Papers and Proceedings* 102, 466–469.

Regan, D. H. (2015), "Explaining Trade Agreements: The Practitioners' Story and the Standard Model," *World Trade Review* 14, 391-417.

Rodrik, D. (2018), "What Do Trade Agreements Really Do?" *Journal of Economic Perspectives* 32, 73-90.

Staiger R. W. and A. O. Sykes (2011), "International Trade, National Treatment and Domestic Regulation," *Journal of Legal Studies* 40, 149-203.

Staiger R. W. and A. O. Sykes (2013), “Non-Violations,” *Journal of International Economic Law* 16, 741-75.

Staiger R. W. and A. O. Sykes (2017), “How Important can the Non-Violation Claim be for the GATT/WTO?" *American Economic Journal: Microeconomics* 9, 149-187.

Staiger R. W. (forthcoming), "Non-Tariff Measures and the WTO," in *The World Trade Organization and Economic Development*, Ben Zissimos (ed.), Cambridge: MIT Press.

Sykes, A. O. (2005), "The Economics of WTO Rules on Subsidies and Countervailing Measures," in P.F.J. McCrory, A.E. Appleton, and M.G. Plummer (eds.), *The World Trade Organization: Legal, Economic, and Political Analysis, Vol. 2*, Springer.