Disputes in International Investment and Trade*

Ralph Ossa† Robert W. Staiger‡ Alan O. Sykes§

January 2020

Abstract

International investment agreements employ dispute settlement procedures that differ markedly from their counterparts in trade agreements along three key dimensions: standing (i.e., the right to file grievances), the nature of the remedy, and the remedial period. In the state-to-state dispute settlement procedures of a typical trade agreement, only governments have standing, while private investors also have standing in the investor-state dispute settlement procedures employed by investment agreements. Trade agreements typically employ tariff retaliation as the remedy for violation of the agreement, while the award of cash damages is the norm in investment disputes. And trade agreements typically provide for only prospective remedies covering harm done subsequent to a ruling, while the damages awarded in investment disputes routinely cover past as well as future harms. We develop parallel models of trade agreements and investment agreements and employ them to study these differences. We argue that the differences can be understood as arising from the fundamentally different problems that trade and investment agreements are designed to solve.

*We thank participants at the 2019 NYU-Columbia Trade Conference for helpful comments. Ralph Ossa gratefully acknowledges funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 819394). The usual disclaimer applies.

†Department of Economics, University of Zurich; and CEPR.

‡Department of Economics, Dartmouth College; and NBER.

§Stanford Law School.

1. Introduction

Firms that supply goods and services to customers in foreign countries are typically protected by two types of international agreements. If they engage in exporting, they are protected by trade agreements, which secure their access to foreign markets. If they engage in foreign direct investment, they are protected by investment agreements, which protect their foreign assets from expropriation and related practices. Increasingly, such trade and investment agreements are combined with other provisions into more comprehensive international economic arrangements. For example, the Comprehensive Economic and Trade Agreement (CETA) between Canada and the EU, the proposed Transatlantic Trade and Investment Partnership (TTIP) between the EU and the US, and the proposed Trans-Pacific Partnership (TPP) between Pacific Rim countries (and its offshoot, TPP11) all offer trade and investment protections.[1]

Dispute settlement pursuant to these agreements has become a subject of much political controversy. The predominant remedy for breach of investment treaty commitments – investor-state dispute settlement (ISDS) with a private right of action for money damages – has come under attack from both the left and the right. In a recent editorial, for example, Senator Elizabeth Warren wrote: "Conservatives who believe in U.S. sovereignty should be outraged that ISDS would shift power from American courts, whose authority is derived from our Constitution, to unaccountable international tribunals. Libertarians should be offended that ISDS effectively would offer a free taxpayer subsidy to countries with weak legal systems. And progressives should oppose ISDS because it would allow big multinationals to weaken labor and environmental rules."[2] An October, 2017 letter to President Trump signed by 230 law and economics professors urged the President to eliminate ISDS from NAFTA, insisting that ISDS "undermines the important roles of our domestic and democratic institutions, threatens domestic sovereignty, and weakens the rule of law."[3] The Trump administration has responded by negotiating the phase-out of ISDS for disputes between Canada and the United States in the new United States Mexico-Canada Agreement (USMCA), which preserves the existing ISDS mechanism for claims involving Mexico only in a few sectors such as telecommunications and oil

[1] In addition to the agreement establishing the World Trade Organization (WTO), there are nearly 300 "regional trade agreements" in force. International investment agreements are even more numerous, with over 2000 bilateral investment treaties in force, and roughly 300 additional investment agreements that are part of larger economic arrangements (usually free trade agreements).

[2] Washington Post, February 25, 2015.

[3] Available at https://www.citizen.org/system/files/case_documents/isds-law-economics-professors-letter-oct-2017_2.pdf.

and gas.[4] Interestingly, much of the resistance to eliminating the ISDS mechanism came from the Mexican government itself, which secured its partial retention over the initial objections of the United States.[5]

The remedy for breach of trade treaty commitments – state-to-state dispute settlement (SSDS) – has also come under attack. On the academic front, critics of the WTO lament the fact that nations can violate treaty obligations yet suffer no formal sanction until such time as the violation is detected, a case is brought and litigated to conclusion, and the violator has had a "reasonable time" to desist from its illegal behavior. This system is said to give violators a "three-year free pass."[6] More recently, United States Trade Representative Robert Lighthizer stated at the 2017 ministerial meeting of the WTO: "The WTO is losing its essential focus on negotiation and becoming a litigation-centered organization. Too often members seem to believe they can gain concessions through lawsuits that they could never get at the negotiating table."[7] And in July, 2018, Lighthizer referred to a recent WTO dispute panel ruling as "the latest example of judicial activism at the WTO seeking to undermine (our unfair trade laws)."[8] Dissatisfaction with WTO dispute rulings has led the United States to block the appointment of judges to vacancies on the WTO Appellate Body for that past several years, creating a situation in which there are now too few confirmed judges to staff the required three-person appellate panels.

All of these controversies arise against a backdrop of intriguing puzzles that the formal economic literature has done little to address. Why do international investment agreements overwhelmingly provide for ISDS rather than SSDS? Why do trade agreements uniformly provide for SSDS and not ISDS?[9] Are the critics of ISDS right that it unwisely burdens national

[4] Other investors can pursue claims against Mexico on a more limited legal basis and only after litigating in Mexican court for up to 30 months.

[5] On Mexico's position in these negotiations, see https://www.lexology.com/library/detail.aspx?g=c42ad0d6-3240-4e24-ac21-5caca954962c. The opposition to ISDS is not limited to the United States. In Europe, the Belgian province of Wallonia temporarily blocked the entire CETA agreement over concerns about its investment provisions. Opposition to these investment provisions was also at the heart of the massive protests in Europe against the CETA and TTIP agreements, which brought hundreds of thousands of Europeans to the streets.

[6] See Davey (2010).

[7] https://www.tralac.org/discussions/article/12560-the-appellate-body-crisis.html.

[8] https://ustr.gov/about-us/policy-offices/press-office/press-releases/2018/july/ustr-robert-lighthizer-statement-wto.

[9] As we noted, some agreements have both trade and investment provisions. In the original NAFTA, for example, the investor-rights provisions allow for both ISDS and SSDS. But the "trade" provisions of NAFTA, involving commitments on trade in goods and services and intellectual property rights, rely exclusively on SSDS, as does the WTO.

sovereignty and undermines sound regulatory policies? Why did Mexico push to retain the ISDS in the USMCA? Might private rights of action in the form of exporter-state dispute settlement (ESDS) make sense in the international trade arena in light of recent attacks on SSDS?

To answer these and related questions, we develop parallel models of trade and investment agreements and employ them to study the relative merits of SSDS and ESDS/ISDS in each setting. Our baseline model of trade agreements mirrors closely that of Maggi and Staiger (2011), where the governments of an importing and an exporting country can make use of a vaguely worded contract written ex ante in the presence of uncertainty about the future state of the world, and where the dispute settlement procedure of the trade agreement involves a court whose mandate is to interpret the contract when invoked ex post in a dispute over whether the trade policy commitments described in the contract have been honored. Our model of investment agreements extends this setting to reflect investment policies and to include an ex-ante investment stage, with foreign investors who at the time of their investment decisions form expectations about the ex-post treatment they will receive from the host-country government under the dispute settlement procedures of the bilateral investment treaty (BIT).

Our analysis focuses on three distinct features of dispute settlement that are often conflated: *standing*, the *nature of the remedy*, and the *remedial period*. *Standing* concerns who has the right to bring cases to a dispute panel. Generally speaking, only member governments have the right to file cases under trade agreements. Under investment agreements, by contrast, private investors also have standing in the vast majority of them.

The *nature of the remedy* concerns the remedial consequences of an adjudicated violation. In trade agreements, adjudicators usually recommend to a violator that it cease and desist from the violation. Sometimes (as in the WTO) the dispute settlement process may also authorize trade sanctions (formally, a "suspension of reciprocal concessions") against violators that refuse to comply with such recommendations. Money awards to complainants are not used as a practical matter (although a violator might pay money to a complaining nation to settle a case). Under investment agreements, by contrast, money awards are routinely available to successful private investors when the agreements grant them standing.

The *remedial period* refers to the period of time that is "covered" by the remedial measure and, in particular, to the question whether any remedy exists for harm done to the complainant prior to the adjudication of a violation. In trade agreements where adjudicators simply recommend that violators cease and desist, the complainant plainly receives no remedy for the harm

done between the commencement of the violation and the point in time at which the violator cures the violation. And where trade retaliation is employed against recalcitrant violators as in the WTO, it tends to be based on the prospective harm from the continued violation, and does not seek to offer any "redress" for retrospective harm. Under investment agreements, by contrast, private investors routinely receive money damages for the entirety of the harm suffered as a consequence of the violation, compensating them for past losses as well as the value of future harms. We refer to the choice of remedial period as the choice between affording only "prospective damages," or in addition providing "retrospective damages."

The first stage of our analysis concerns *standing*. We argue that the key difference across governments and private litigants is that exporters and private investors would invoke the court more aggressively than would governments on their behalf. Our analysis of the issue of standing then proceeds by asking whether there can be conditions under which this greater litigiousness is desirable.

We find that if treaties are concerned primarily with *market access commitments*, which we argue are the central focus of trade treaties and which we therefore adopt as the focus of trade agreements in all of our analysis below, then SSDS dominates ISDS because it eliminates socially undesirable litigation by private actors. If the adjudicative process is highly inaccurate, SSDS may itself be counterproductive and render inutile trade agreements that allow courts to interpret the contract: but in no circumstance does standing for private firms improve upon the situation. Intuitively, this follows from two observations. First, market access commitments are fundamentally government-to-government policy commitments that address an inefficiency that can be traced to an international policy externality, and so the optimal choice of standing will maximize the surplus that the agreement can create for the two governments. And second, under SSDS the governments themselves would adopt behavior which leads to too much litigation, owing to the negative external effects that are not internalized in their decision to litigate; and so granting standing to private actors that would add to this (overly-) litigious behavior will lower the joint surplus obtained by the governments under the agreement.

In contrast, if treaties are concerned primarily with *government commitment problems*, which we argue are the central focus of investment treaties and which we therefore adopt as the focus of investment agreements in all of our analysis below, then standing for private actors may well be desirable as a way to overcome imperfections in the extent to which home governments represent the interests of their domestic firms operating abroad. Factors that

bear on the desirability of standing for private actors include the accuracy of the adjudicative process, the probability that a host country will take an action that can be alleged to violate treaty commitments and the degree of ex-post efficiency associated with that action, and the litigation cost borne by the host country in the event of a formal dispute. Intuitively, these findings follow because the commitment problem at the heart of investment agreements is a government-to-investor policy commitment problem that creates a host-country inefficiency, and so the optimal choice of standing will maximize the surplus that the agreement can create for the host government. And from the host government's perspective, the foreign government can be either too litigious or not litigious enough, and in the latter case the desirability of standing for private actors can arise.

Hence, unlike in the context of trade agreements, our analysis offers qualified support for the inclusion of private standing in investment agreements, though by no means suggests that it is universally optimal in this context. Along the way, we also derive conditions on the minimal level of adjudicative accuracy under which investment agreements do a better job at solving commitment problems than up-front investment incentives by host countries.

The second stage of our analysis concerns the *nature of the remedy* – the choice between retaliatory sanctions on the one hand, and money damages on the other. In developing the analysis, we take as a starting point the fact that trade agreements limit standing to governments, while investment agreements afford standing to investors, a situation that receives qualified support from our discussion of standing. The key tradeoff that we exploit is the observation that sanctions tend to create significant deadweight costs and are more costly than money damages, while damages are more difficult to compute. Plausibly, the computation of damages for a single investor injured by a host government is much more straightforward, and likely to be accurate, than the computation of damages for a government whose exporters have suffered harm. This fact will weigh in favor of damages under investment agreements and against damages under trade agreements but is not the only relevant consideration – one must also take into account the likelihood that errors in court rulings will lead to inefficient choices by defendants under each regime.

The third stage of our analysis adds consideration of the *remedial period* in the dispute process as we extend our baseline models of trade and investment agreements to consider the choice between prospective damages only on the one hand, and the addition of retrospective damages on the other. In developing the analysis of this choice, we rely on the fact that

the formal enforcement mechanism in trade treaties generally consists of costly retaliatory trade sanctions, while the enforcement mechanism in investment treaties generally involves (less costly) monetary transfers. We then find that the retrospective remedy will be preferable if the retrospective portion of the harm suffered as a consequence of treaty violations is high enough and any inefficiencies associated with the remedy are small enough – as seems reasonable for the case of investment disputes, where the pre-ruling harm is usually attributable to policy actions that diminish or destroy the returns to a sunk investment with a finite lifespan and where the remedy takes the form of monetary transfers. In contrast, we find that the prospective remedy alone is preferable when the retrospective portion of the harm suffered as a consequence of treaty violations is low enough and/or the efficiency costs of the remedy are high enough – as seems reasonable for the case of trade disputes, where the pre-ruling harm is attributable mostly to the costs of the delay in securing market access and where the remedy takes the form of retaliatory trade sanctions.

Our analysis of the remedial period in the dispute process thus provides support for the retrospective damages found in BITs. But perhaps surprisingly, our analysis also provides some support for the prospective damages approach taken by the WTO and trade agreements more broadly, in that, while not perfect, allowing violators to take a "three-year free pass" may be the better option if the alternative involves retrospective damages paid out in the form of highly inefficient retaliatory trade sanctions.

Our paper is related to several literatures. A first literature focuses on dispute settlement procedures in trade agreements (see Park, 2016, for a recent survey); within this literature, our paper is most closely related methodologically to Maggi and Staiger (2011), who provide the basic model of dispute settlement in trade agreements on which we build. A second and more recent literature focuses on dispute settlement procedures in BITs (see, for example Konrad, 2017, Horn, 2018, Stahler, 2018, Horn and Tangeras, 2019, Janeba, 2019 and Kohler and Stahler, 2019). Our use of the Maggi and Staiger (2011) framework makes us methodologically distinct from these papers. Relative to all of this literature, our paper ventures into largely uncharted territory: to the best of our knowledge, we are the first to formally explore the choice of standing, the nature of the remedy, and the remedial period together in either trade or investment agreements and, by considering these features in both trade and investment agreements, the first to offer a formal comparative analysis of the core design features of dispute

settlement procedures across these two setting.[10]

The remainder of the paper proceeds as follows. Section 2 provides additional institutional detail on trade and investment agreements and their dispute settlement systems. Section 3 contains our analysis of the standing issue, while section 4 considers the nature of the remedy and section 5 focuses on the remedial period. Section 6 concludes.

2. Trade and Investment Disputes in Practice

Important heterogeneities exist within and among trade and investment agreements and their approaches to dispute settlement, and it is not our intent here to provide a comprehensive survey of these heterogeneities. Instead, we focus on "typical" characteristics of trade and investment agreements with reference to a few illustrative examples.

2.1. Trade Agreements

Modern international trade agreements date back to the beginning of Reciprocal Trade Agreements program in the 1930s, which eventually led to the creation of the General Agreement on Tariffs and Trade (GATT) in 1947. The key substantive obligations under trade agreements such as GATT are tariff commitments, along with restrictions on other border measures (such as quotas) that can impede market access, and constraints on various domestic policies (e.g., discriminatory taxes and regulations, subsidies) that can undermine market access commitments associated with tariff reduction.

Under GATT, formal sanctions for violations did not arise in practice, but disputants could agree to allow arbitral panels to adjudicate the merits of a case and issue a report as to whether a violation had occurred. Over time, the membership became dissatisfied with this system and, with the creation of the WTO (which subsumed GATT), established a new system whereby complaining nations can obtain an arbitral panel and secure a final ruling (including a right of appeal), accompanied by a formal recommendation to a violator to cease the violation within a "reasonable time." If a violator fails to do so, the complaining nation may reach an agreement with the violator for alternative compensation. Failing such agreement, the complainant may impose retaliatory trade measures (a "suspension of concessions") in an amount "equivalent"

[10] Horn (2018) considers the implications of switching from ISDS to SSDS in the context of BITs in a model similar to Horn and Tangeras (2019), but he does not optimize other features of the agreement and does not consider trade agreements.

to the level of harm done by the violation, an issue subject to arbitration if requested by the violator. In practice, arbitrators judge "equivalence" in relation to the prospective harm from the ongoing violation. (Davey 2010). Scholars have debated whether this system should be viewed as a "property rule" or liability rule" (Schwartz & Sykes 2002), and we take no position on the issue. But three observations are beyond dispute – only member state governments have standing to file cases, the system relies on negotiated compensation and retaliatory trade measures rather than monetary damage awards as its sole formal "remedy," and at no time in the history of WTO/GATT has there been any formal remedy for harm done prior to the adjudication of a violation.

The dispute settlement systems under the hundreds of free trade agreements now in force vary somewhat, but many share core features with the WTO system. Under NAFTA Chapter 20 (preserved in the USMCA), for example, only member governments can bring complaints in relation to the trade provisions. A complaining party can seek an arbitral panel, which will rule on the existence of a violation. If one is found, the remedy "wherever possible" is removal of the offending measure, or negotiated trade compensation. Failing successful negotiation in that regard, the complaining party again has the option to suspend "benefits of equivalent effect" to the violation. There is no monetary remedy, and the concept of "equivalence" is forward looking in relation to the harm done by the ongoing violation until such time as it is eliminated or the case is settled. Broadly similar provisions are to be found in other U.S. free trade agreements.

2.2. Investment Agreements

Investment agreements address two broad classes of issues. Foremost, they protect existing investors against certain measures by host countries that impair the value of their established investments. These include expropriation and equivalent actions without adequate compensation, discrimination in favor of host country or third-country investors, and in many agreements a guarantee of "fair and equitable treatment" that addresses behavior by host governments involving fraud, deception and denial of due process. In a considerably smaller number of investment treaties (only about six percent of all BITs surveyed by UNCTAD [2018]), investment agreements also secure non-discriminatory "market access" for foreign investors. The core obligation here is "national treatment" (non-discrimination relative to domestic investors) in the "establishment" of investments. Investment agreements that contain this provision enable foreign investors to compete on equal terms with domestic investors for new investment oppor-

tunities (with some exceptions, such as national security concerns). All of the existing United States BITs and free trade agreements with investment provisions include national treatment for establishment of investments.

SSDS in some form can be found in virtually all international investment agreements. But in addition, ISDS is included in 95% of BITs currently in force according to UNCTAD (2018), and in all international investment agreements involving the United States except for its limited role in the new USMCA as noted earlier. The details vary as to the exact preconditions for invoking ISDS, but the basic structure allows investors who believe that a treaty commitment has been violated, and who have not secured adequate redress through consultations or litigation in the host country, to bring a case to international arbitration before neutral arbitrators. The most common arbitrations occur at the International Center for the Settlement of Investment Disputes (ICSID) at the World Bank, or in *ad hoc* arbitration pursuant to the UNCITRAL arbitration rules. The arbitrators determine whether the dispute falls within their jurisdiction under the terms of the treaty that is invoked, and if so adjudicate the merits of claims regarding treaty violations. When a violation is found, the arbitrators proceed to assess damages (the "quantum"), and issue an award directing the host country to compensate the complaining investor for the violation. In cases of expropriation of property, they may offer the host country an opportunity to reduce damages by making restitution of the property. The general principle governing damages is that the investor should be compensated for all past and future injury due to the violation. Most awards are paid voluntarily by the host country, and their enforcement is facilitated by the "New York Convention" which requires signatories to treat the awards as the equivalent of awards by their own domestic courts.

3. Standing

In this section we focus exclusively on issues of standing, and thus the question of whether to include state-to-state dispute settlement procedures (SSDS) or exporter-to-state/investor-to-state dispute settlement procedures (ESDS/ISDS) in an optimally designed trade/investment agreement. We first consider the standing issue when the underlying problem being solved by the agreement relates to market access/terms of trade issues. Our formal results are derived in the setting of a trade agreement, reflecting the position that market access/terms of trade issues are the central issues of concern in trade agreements, but we also comment on how our results extend to market access issues that are handled in international investment agreements. And

we then address the standing issue when the underlying problem involves making commitments to foreign firms that must make sunk investments to serve a domestic market. There our formal results are derived in the setting of an investment agreement, reflecting the dominance of such issues for BITs; and there we comment on how our results extend to commitment issues that are handled in trade agreements.

To isolate the standing issue, we make two simplifying assumptions in this section. First, we assume that damage payments are not part of the court's ruling - when a case is filed and the court sides with the complainant, the defendant has no choice but to "cease and desist" whatever policy is found to be illegal. Second, we assume that compliance with any such ruling is instantaneous so that there is no pre-compliance harm suffered by exporters or investors for which additional remedies might be desirable. We relax these restrictions in sections 4 and 5, where we analyze the optimal nature of the remedy (cash versus retaliation) and the optimal remedial period (prospective versus retrospective).

3.1. Disputes over Market Access in Trade Agreements

We begin by considering the standing issue in a trade agreement building on the model of Maggi and Staiger (2011). We focus on a single industry in which Home is an importer and Foreign is an exporter. Home and Foreign have a trade agreement which has a dispute settlement body (DSB) to resolve disputes. Home has to make a binary policy choice $\tau \in \{FT, P\}$ (free trade or protection) and Foreign has to decide whether to file a complaint with the DSB (there are no export policy instruments). We distinguish between a system of SSDS in which the filing choice is made by the foreign government and a system of ESDS in which the filing choice is made by the foreign exporting industry. There are $s \equiv (s_1, s_2, \ldots, s_N)$ binary states of the world, such as "there is/is not an import surge," and we let $p(s)$ denote the probability that state s occurs.[11]

The importer government gains from protection and both the exporter government and the exporting industry suffer from protection in all states of the world. We denote the home government's payoff from choosing policy τ in state s as $\omega(\tau; s)$ and assume that $\gamma_G(s) \equiv \omega(P; s) - \omega(FT; s) > 0$ for all s. Similarly, we denote the foreign agent's payoff as $\omega_f^*(\tau; s)$,

[11] We follow Maggi and Staiger (2011) and Staiger and Sykes (2017) and abstract from the possibility of negotiated settlements to a dispute (i.e. negotiations between the two governments over the importer government's policy choice after the state s is realized). See Maggi and Staiger (2018) for an analysis of trade disputes that features the possibility of settlement.

where $f \in \{G^*, E^*\}$ depending on whether we consider the government's or the exporting industry's payoff, and assume $\gamma_f^*(s) \equiv \omega_f^*(P;s) - \omega_f^*(FT;s) < 0$ for all s. These assumptions can be given a terms-of-trade interpretation and would hold in any standard trade model. Moreover, $\omega(\tau;s)$ and $\omega_{G^*}^*(\tau;s)$ can also capture distributional/political economy considerations such as putting a higher weight on producer surplus than consumer surplus or tariff revenue.

Free trade is the "first best" policy in all states $s \in \sigma^{FT}$ and protection is the "first-best" policy in all states $s \in \sigma^P$. We define the first-best policy as the policy that maximizes the governments' joint payoff $\Omega(\tau;s) \equiv \omega(\tau;s) + \omega^*(\tau;s)$ and let $\Gamma(s) \equiv \Omega(P;s) - \Omega(FT;s) = \gamma_G(s) + \gamma_{G^*}(s)$ denote the governments' joint (positive or negative) gain from protection so that $s \in \sigma^{FT}$ if $\Gamma(s) \leq 0$ and $s \in \sigma^P$ if $\Gamma(s) > 0$. In the absence of a trade agreement, the importer government would choose $\tau = P$ in all states of the world. This noncooperative policy choice would correspond to the first best for $s \in \sigma^P$, but it would differ from the first best for $s \in \sigma^{FT}$, giving rise to the possibility that the two governments could do better under a trade agreement. We assume that the realized state s is observed by all agents including the DSB, but that the DSB does not observe Γ and hence payoff levels are not verifiable. This means that the first-best outcome cannot be trivially achieved with a contract that requires FT if and only if $\Gamma \leq 0$.

Following Maggi and Staiger (2011), we assume as well that it is prohibitively costly to describe precisely all the relevant state variables $(s_1, s_2, \ldots, s_N)$ that would be necessary to write a complete contingent contract covering the policy τ. We focus instead on what Maggi and Staiger call a "vague contract" that takes the form "$\tau = P$ allowed if and only if ν," where ν is a vague sentence such as "there is serious injury to the domestic industry due to increased imports." This off-the-shelf language makes the vague contract essentially costless to write, and following Maggi and Staiger we assume that it specifies the first-best policy choice in those states of the world where its meaning is unambiguous. But the meaning of this contract is ambiguous in some states of the world, and it is in such states that a dispute over the setting of τ may arise in our model.

Specifically, we assume that governments have given the DSB a mandate to serve an "interpretive" role: if invoked, the DSB observes an unbiased but noisy signal of Γ, which can be thought of as the outcome of an independent investigation in which the DSB "interprets" the contract. The DSB then issues a "cease-and-desist" ruling $\tau^{DSB} = FT$ if its signal indicates $\Gamma \leq 0$, and it issues the ruling $\tau^{DSB} = P$ if its signal indicates $\Gamma > 0$. The DSB ruling can there-

fore be thought of as simply a policy determination that maximizes the expected joint payoff of the governments given the DSB signal.[12] We assume that the ruling is automatically enforced, and we denote the probability that the DSB issues the "wrong" ruling by $qk(s) \in (0, 1/2)$ where $k(s) \in (0, 1/2)$ for all s and $q \in (0, 1)$ parameterizes the (inverse) quality of the court.[13] Invoking the court is costly and we write the litigation costs incurred by Home's government and Foreign's complainant as $c(s) = c\epsilon(s)$ and $c_f^*(s) = c_f^*\epsilon^*(s)$, where $\epsilon(s)$ and $\epsilon^*(s)$ are positive state-dependant constant and c and c_f^* parameterize the litigation costs.

The timing of events is as summarized in Figure 1. First, the state of the world realizes and either $s \in \sigma^{FT}$ or $s \in \sigma^P$. Then, the home government moves and makes its binary policy choice $\tau \in \{FT, P\}$. If the home government chooses $\tau = FT$, the foreign complainant never has an incentive to invoke the DSB so that free-trade prevails. If the home government instead chooses $\tau = P$, the foreign complainant has to weigh its options and may or may not invoke the DSB. If the foreign complainant does not invoke the DSB, protection prevails. If the foreign complainant instead invokes the DSB, the DSB moves and issues its ruling $\tau^{DSB} = FT$ or $\tau^{DSB} = P$ which is then adopted instantaneously.

This game is straightforward to solve by backwards induction. Consider first the foreign complainant's filing behavior. The complainant files a complaint if and only if $\tau = P$ and the expected benefit to the complainant of filing exceeds the complainant's cost of filing, that is

$$\Pr(\text{DSB ruling is } FT \mid s) \times |\gamma_f^*(s)| > c_f^*(s). \tag{3.1}$$

Condition (3.1) is the "filing" condition for the agent with standing in the exporting country to invoke the DSB in response to a policy choice by the importer government of $\tau = P$.

[12] We follow Maggi and Staiger (2011) in assuming that the DSB seeks to maximize the governments' joint payoff – and therefore attempts to complete the contract as the governments would have done ex ante – and Maggi and Staiger provide support for this assumption at a broad level. Nevertheless, as Maggi and Staiger discuss, this assumption glosses over a number of features of real-world trade agreements. In particular, rather than undertaking a broad assessment of the international efficiency of a particular policy measure, a typical DSB ruling would in reality concern itself with the more narrow question of whether the measure was in compliance with various explicit (but vaguely worded) commitments contained in the contract (e.g., national treatment, or MFN) which themselves can be interpreted as attributes of internationally efficient policy intervention. Like the model of Maggi and Staiger, our model does not include this extra layer of mapping between, on the one hand, vague contractual commitments, and on the other hand, the international efficiency properties of measures which conform to these commitments. We also abstract from the possibility that the court might be an active player in the game and attempt to extract information from the behavior of the disputants (see Maggi and Staiger, 2011, for a discussion of the limited possibilities the court might have to engage in such behavior).

[13] Maggi and Staiger (2011) derive conditions under which it is optimal for the governments to write a vague contract and install a court with a mandate to interpret the contract if invoked. We take these two institutional features as given so that we may focus on other dimensions of the design of dispute settlement procedures. See also note 15.

Next consider the importer government's policy choice. This government chooses $\tau = P$ if either (3.1) fails – because then the importer government can set $\tau = P$ without triggering a dispute – or if (3.1) holds and the expected benefit to the importer government from trade protection exceeds the cost to the importer government of a dispute:

$$\Pr(\text{DSB ruling is } P \mid s) \times \gamma_G(s) > c(s). \tag{3.2}$$

We can now derive the equilibrium actions for each state s. For simplicity, in what follows we assume that the states where the vague contract is unambiguous are measure zero, so we can focus only on states where the court if invoked must interpret the contract.[14] Defining the thresholds $\mu_1^{FT}(s) \equiv \frac{c(s)}{\gamma_G(s)}$, $\mu_{2,f}^{FT}(s) \equiv 1 - \frac{c_f^*(s)}{|\gamma_f^*(s)|}$, $\mu_{1,f}^{P}(s) \equiv \frac{c_f^*(s)}{|\gamma_f^*(s)|}$, and $\mu_2^P(s) \equiv 1 - \frac{c(s)}{\gamma_G(s)}$, and noting that $\mu_1^{FT}(s) < \mu_{2,f}^{FT}(s)$ and $\mu_{1,f}^{P}(s) < \mu_2^P(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (3.1) and (3.2) immediately imply the following result:

Lemma 1. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\mu_1^{FT}(s) < \mu_{2,f}^{FT}(s)$ and $\mu_{1,f}^{P}(s) < \mu_2^P(s)$, equilibrium actions can be characterized as follows:*

1. *In states $s \in \sigma^{FT}$: If DSB quality is high in the sense that $qk(s) < \mu_1^{FT}(s)$, we have $\tau = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\mu_1^{FT}(s), \mu_{2,f}^{FT}(s)\right]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu_{2,f}^{FT}(s)$, we have $\tau = P$ and no dispute.*

2. *In states $s \in \sigma^{P}$: If DSB quality is high in the sense that $qk(s) < \mu_{1,f}^{P}(s)$, we have $\tau = P$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\mu_{1,f}^{P}(s), \mu_2^P(s)\right]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu_2^P(s)$, we have $\tau = FT$ and no dispute.*

This result is visualized in Figure 2a. Notice that the court has its best impact off-equilibrium, when due to its high accuracy it induces both the importer government and the potential foreign complainant to behave efficiently in order to avoid a dispute. Where a dispute arises in equilibrium, there must be opportunistic behavior on the part of either the importer

[14] This is without loss of generality, because under our assumptions in states where the vague contract is unambiguous the importer government would make the first best policy choice and their would be no filing by the foreign complainant, and hence nothing of consequence for any of the results we emphasize.

government (if the importer government is exploiting the incompleteness of the contract and the inaccuracy of the DSB and trying to get away with protection when free trade is efficient) or the foreign complainant (if the foreign complainant is exploiting the incompleteness of the contract and the inaccuracy of the DSB and trying to force free trade when protection is efficient). And finally, if the DSB is inaccurate enough (i.e., for $qk(s) > \mu_{2,f}^{FT}(s)$ in states $s \in \sigma^{FT}$ and for $qk(s) > \mu_2^P(s)$ in states $s \in \sigma^P$) its beneficial off-equilibrium impact will erode, and such opportunistic behavior can arise while the DSB sits on the sideline.

We can now write down the expected efficiency loss, relative to the first-best outcome, that is associated with standing choice $f \in \{G^*, E^*\}$ in combination with the vague contract and interpretive court mandate, a combination of design features that we denote by V_f and refer to as the V_f *institution*. Denoting this efficiency loss by $L(V_f)$ and defining the sets $\hat{\sigma}_1^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) < \mu_1^{FT}(s)\right\}$, $\hat{\sigma}_{2,f}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) \in \left[\mu_1^{FT}(s), \mu_{2,f}^{FT}(s)\right]\right\}$, and $\hat{\sigma}_{3,f}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) > \mu_{2,f}^{FT}(s)\right\}$, as well as $\hat{\sigma}_{1,f}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) < \mu_{1,f}^{P}(s)\right\}$, $\hat{\sigma}_{2,f}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) \in \left[\mu_{1,f}^{P}(s), \mu_2^{P}(s)\right]\right\}$, and, $\hat{\sigma}_3^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) > \mu_2^{P}(s)\right\}$, we can write:

$$\begin{aligned} L(V_f) \;=\; & \sum_{s \in \left\{\hat{\sigma}_{2,f}^{FT} \cup \hat{\sigma}_{2,f}^{P}\right\}} p(s)\, qk(s)\, |\Gamma(s)| + \\ & \sum_{s \in \left\{\hat{\sigma}_{2,f}^{FT} \cup \hat{\sigma}_{2,f}^{P}\right\}} p(s) \left[c(s) + c_f^*(s)\right] + \\ & \sum_{s \in \left\{\hat{\sigma}_{3,f}^{FT} \cup \hat{\sigma}_{3}^{P}\right\}} p(s) |\Gamma(s)|. \end{aligned} \tag{3.3}$$

Each line of expression (3.3) captures a distinct source of inefficiency arising under the V_f institution. The first line captures the loss associated with DSB error, and it is the product of the probability that state s occurs, $p(s)$, the probability that the DSB makes a mistake, $qk(s)$, and the efficiency loss associated with that mistake, $|\Gamma(s)|$, summed over all states in which the DSB is invoked, $s \in \left\{\hat{\sigma}_{2,f}^{FT} \cup \hat{\sigma}_{2,f}^{P}\right\}$. The second line captures the efficiency loss arising from the cost of a dispute, and it is the product of the probability that state s occurs, $p(s)$, and the joint cost of a dispute in state s, $c(s) + c_f^*(s)$, again summed over all states in which the DSB is invoked, $s \in \left\{\hat{\sigma}_{2,f}^{FT} \cup \hat{\sigma}_{2,f}^{P}\right\}$. The third line is the efficiency loss arising from distorted choices made "in the shadow of the court", and it is the product of the probability that state s occurs, $p(s)$, and the efficiency loss from getting the inefficient outcome in state s, $|\Gamma(s)|$, summed over all states in which the DSB quality is so poor that the inefficient policy choice prevails without any dispute, $s \in \left\{\hat{\sigma}_{3,f}^{FT} \cup \hat{\sigma}_{3}^{P}\right\}$.

SSDS versus ESDS We now evaluate the desirability of adopting SSDS versus ESDS. Under SSDS (the V_{G^*} institution), the foreign government has standing, in the sense that it alone has the right to file a dispute with the DSB. Under ESDS (the V_{E^*} institution), the foreign exporting industry/firm has standing, in the sense that it alone has the right to file a dispute with the DSB. In reality, the relevant thought experiment would more likely be to compare the institution with SSDS to a counterfactual alternative institution that features *both* SSDS and ESDS where both the foreign government and the foreign exporter have standing: for simplicity, we choose to proceed formally in this more parsimonious way, and with our formal arguments in hand to then draw observations relevant to the more realistic possibility of ESDS as an addition to (rather than a replacement for) SSDS.

Notice that the question of optimal standing is only relevant for intermediate levels of DSB quality. If the quality is so high that $qk(s) < \min\left\{\mu_1^{FT}(s), \mu_{1,f}^{P}(s)\right\}$ for all s, the trade agreement always yields the efficient outcome regardless of standing (though the critical quality level is itself a function of the choice of standing f), with protection always occurring with impunity when it should (in σ^P) and never occurring when it should not (in σ^{FT}). If the quality is so low that $qk(s) > \max\left\{\mu_{2,f}^{FT}(s), \mu_2^{P}(s)\right\}$ for all s, the trade agreement always yields the inefficient outcome regardless of standing, with protection never occurring when it should (in σ^P) and always occurring with impunity when it should not (in σ^{FT}). In this latter case, the governments would even be better off without any trade agreement, since the noncooperative equilibrium at least delivers the efficient outcome in states $s \in \sigma^P$, and we rule out this case by assumption henceforth.[15]

According to (3.3), the relative merits of SSDS versus ESDS can be evaluated once we specify the complainant's cost of filing and payoff from winning in court under each choice of standing, that is, when only the foreign government can be the complainant as under SSDS, and when only the foreign exporting industry can be the complainant as under ESDS. This is because $c_f^*(s)$ enters equation (3.3) directly and $c_f^*(s)$ and $\left|\gamma_f^*(s)\right|$ enter it indirectly by determining the sets $\hat{\sigma}_{2,f}^{FT}$, $\hat{\sigma}_{2,f}^{P}$, and $\hat{\sigma}_{3,f}^{FT}$ through the conditions (3.1) and (3.2). Recall that

[15] We are ignoring here an issue highlighted in the analysis of Maggi and Staiger (2011) who treat the contract and court mandate as an endogenous institutional choice. They show that with low court quality the vague contract with an interpretive court is no longer the optimal institution; instead, it becomes optimal to reign in the court's "activist" mandate to settle disputes as they arise (and to possibly change the form of the contract), and the resulting "passive" court's role becomes entirely off-equilibrium. For our purposes here, however, the point we emphasize would survive endogenizing the institutional choice in this way: our point is simply that with low enough court quality, it would not be optimal to introduce a trade agreement with an activist court where the issue of standing would arise.

$c_f^*(s) = c_f^* \epsilon^*(s)$ is the cost of filing and $\gamma_f^*(s)$ is the loss associated with protection relative to free trade of foreign complainant $f = \{G^*, E^*\}$.

We assume that the cost of filing for the foreign exporting industry is the same as the cost of filing for the foreign government, namely

$$c_{G^*}^* = c_{E^*}^* \equiv c^* \qquad \text{(Assumption 1)}$$

implying $c_{G^*}^*(s) = c_{E^*}^*(s) \equiv c^*(s)$ for all s. Our key assumption, under which the choice of standing is consequential for the efficiency properties of the trade agreement, is that the loss from protection suffered by the foreign exporting industry is *greater* than the loss suffered by the foreign government, or

$$|\gamma_{G^*}^*(s)| < |\gamma_{E^*}^*(s)|. \qquad \text{(Assumption 2)}$$

Note that Assumption 2 would hold in any standard trade model since Home's tariff reduces prices in Foreign, which harms producers but benefits consumers, and Foreign's government also takes consumer interests into account. But it could also capture, for example, a Ricardo-Viner logic whereby home protection will create losers in the foreign country (those interests tied to the foreign export industry, with losses $\gamma_{E^*}^*(s) < 0$) but also winners (those interests tied to the foreign import-competing industry), implying Assumption 2 provided only that the foreign government places some weight on the gains enjoyed by the foreign import-competing industry when the home government raises trade barriers against foreign exporters. More generally, Assumption 2 reflects the fact that governments can maintain a "political filter" when deciding what cases to bring under SSDS, and that political filter is lost under ESDS. In effect, then, by Assumption 1 and Assumption 2 adopting ESDS rather than SSDS amounts to the foreign government delegating filing decisions to a more aggressive filer than itself.[16] The private filing decision of foreign exporters under ESDS takes account of the private benefits of litigation to the foreign exporters but ignores the costs to other actors in their country.

[16] We are abstracting here from a potentially important free-rider issue that could arise under ESDS, namely, the same free-rider issue that can arise in an industry lobbying setting. If the firms in an industry cannot overcome this free-rider issue, filing might not occur under ESDS even though it would be in the collective interest of the firms in the industry to file. This could be captured in our model with the possibility that $|\gamma_{E^*}^*(s)| < |\gamma_{G^*}^*(s)|$, just the opposite of what we assume in Assumption 2. However, as we noted at the outset of this subsection, while we formally evaluate the desirability of adopting ESDS *rather* than SSDS, in reality, the relevant thought experiment would more likely be to add ESDS to SSDS so that both the foreign government and the foreign exporter have standing. And in that case, the SSDS would handle filings for situations where the free-rider issue resulted in $|\gamma_{E^*}^*(s)| < |\gamma_{G^*}^*(s)|$, and the ESDS would only be relevant for cases where Assumption 2 applies. Hence, we are abstracting from this free-rider issue without loss of generality.

The comparison between the V_{G^*} and V_{E^*} institutions can be made by comparing their respective losses relative to the first best using (3.3). To facilitate this comparison, we define two sets that embody the two key changes that would occur according to Assumption 1 and Assumption 2 if standing were taken from the foreign government and given instead to the foreign exporting industry, and illustrate them in Figure 2b. The first set is $\Omega_S^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) \in \left[\mu_{2,G^*}^{FT}(s), \mu_{2,E^*}^{FT}(s)\right]\right\}$: for $s \in \Omega_S^{FT}$ the importer government would choose the inefficient $\tau = P$ with impunity under the V_{G^*} institution but a court filing would occur under V_{E^*}. The second set is $\Omega_S^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) \in \left[\mu_{2,E^*}^{P}(s), \mu_{2,G^*}^{P}(s)\right]\right\}$: for $s \in \Omega_S^{P}$ the exporter government would allow the efficient choice $\tau = P$ to go unchallenged under the V_{G^*} institution but a court filing would occur under V_{E^*}. Defining $\Delta_{E^*,G^*} \equiv L(V_{E^*}) - L(V_{G^*})$, we can then write:

$$\begin{aligned} \Delta_{E^*,G^*} &= \sum_{s\in\Omega_S^P} p(s)\left[qk(s)\left|\Gamma(s)\right| + c(s) + c^*(s)\right] + \\ &\quad \sum_{s\in\Omega_S^{FT}} p(s)\left\{\left[qk(s)\left|\Gamma(s)\right| + c(s) + c^*(s)\right] - \left|\Gamma(s)\right|\right\}. \end{aligned} \tag{3.4}$$

This expression highlights the costs and benefits associated with the additional litigation arising under ESDS relative to SSDS. The first line of (3.4) is clearly positive and captures that the extra litigation is undesirable in states $s \in \Omega_S^P$ since it then challenges an efficient policy thus introducing nothing but court error and litigation costs (note that $qk(s)\left|\Gamma(s)\right|$ is the expected loss associated with court error and $c(s) + c^*(s)$ is the joint litigation cost under Assumption 1). The second line of (3.4) is also positive and captures that the extra litigation is also undesirable in states $s \in \Omega_S^{FT}$ but the sign is less obvious this time. This is because the extra litigation now challenges an inefficient policy so that the loss associated with court error and litigation costs is counterbalanced by an efficiency gain $\left|\Gamma(s)\right|$. However, it is easy to show that the second line is also positive using the facts that $qk(s) > \mu_{2,G^*}^{FT}$ and $\Gamma(s) < 0$ in $s \in \Omega_S^{FT}$.[17] Intuitively, for $s \in \Omega_S^{FT}$ the exporter *government* does not see a filing as worth the dispute cost while the importer government never benefits from a filing, and the fact that foreign exporters would nevertheless choose to file simply reduces the value of the agreement to the two governments. Hence, $\Delta_{E^*,G^*} > 0$ provided that at least one of the sets Ω_S^P or Ω_S^{FT} is non-empty.

[17] In particular, $qk(s) > \mu_{2,G^*}^{FT}$ implies $c^*(s) > \left[1 - qk(s)\right]\left|\gamma_{G^*}^*(s)\right|$ and $\Gamma(s) < 0$ implies $\gamma_G(s) + \gamma_{G^*}^*(s) < 0$ and thus $\left|\Gamma(s)\right| = -\left[\gamma_G(s) + \gamma_{G^*}^*(s)\right]$ so that we can write $\sum_{s\in\Delta_S^{FT}} p(s)\left\{-\left[1 - qk(s)\right]\left|\Gamma(s)\right| + c(s) + c^*(s)\right\} = \sum_{s\in\Omega_S^{FT}} p(s)\left\{\left[1 - qk(s)\right]\left[\gamma_G(s) + \gamma_{G^*}^*(s)\right] + c(s) + c^*(s)\right\} > \sum_{s\in\Omega_S^{FT}} p(s)\left\{\left[1 - qk(s)\right]\gamma_{G^*}^*(s) + c^*(s)\right\} > 0$.

The two governments would therefore choose to include SSDS rather than ESDS in their trade agreement. Intuitively, this is because under SSDS the governments themselves would adopt behavior which leads to too much litigation, owing to the negative external effects that are not internalized in their decision to litigate; and so granting standing to private actors that would add to this (overly-) litigious behavior will lower the joint surplus obtained by the governments under the agreement. And it is immediate that including *both* SSDS and ESDS would be outcome equivalent to including ESDS *instead of* SSDS. Hence, if given a choice between including both SSDS and ESDS in a trade agreement or including just SSDS, the two government would choose the latter option.

We summarize with:

Proposition 1. *Governments, but not their exporters, should have standing to bring disputes in an optimally designed trade agreement. That is, an optimally designed trade agreement should include SSDS, but not ESDS.*

Standing for market access disputes more generally While we have thus far analyzed market access issues in the context of trade agreements, similar issues could also arise in the context of BITs, given the close relationship between exporting and (horizontal) foreign direct investment (FDI). Exporting and FDI are typically viewed as two alternative ways of serving a foreign market, between which firms choose based on a proximity-concentration trade-off (Brainard, 1997; Helpman, Melitz, and Yeaple, 2004). Exporting has the advantage that it allows firms to concentrate production in one location, while FDI has the advantage that it allows firms to avoid trade costs, so that the optimal mode of accessing a foreign market is determined by the relative importance of plant-level economies of scale and trade costs.

In a companion paper (Ossa, Staiger, and Sykes, 2020), we show that countries therefore have a terms-of-trade motive for restricting the market access of foreign multinationals just as they have a terms-of-trade motive for restricting the market access of foreign exporters. However, this motive is weaker in the context of FDI than in the context of exporting so that market access considerations should play less of a role in BITs than in trade agreements. Qualitatively, restricting the market access for foreign multinationals is the same as restricting the market access for foreign exporters in that both reduce the demand for foreign products and thus improve the terms-of-trade. Quantitatively, however, restricting the market access of foreign multinationals is less effective than restricting the market access of foreign exporters,

since local affiliates of foreign multinationals typically also employ local factors which weakens the ability to shift costs to foreign countries.

In light of this, we view the results of this section as applying to market access disputes more generally, whether they arise in trade agreements or in investment agreements: with regard to market access/terms-of-trade issues, only governments should have standing to bring disputes in an optimally designed trade or investment agreement. That is, for the purpose of settling market access disputes, an optimally designed trade agreement should include SSDS, but not ESDS, while an optimally designed investment treaty should include SSDS, but not ISDS.

3.2. Disputes over Commitments to Investors in Investment Agreements

We next consider the issue of standing in investment agreements, and thus the question of whether they should include SSDS or ISDS. For simplicity, we assume for now that BITs are only concerned with the home (now host) government making policy commitments to foreign investors (i.e., we abstract from any market access issues associated with foreign investors). To this end, we make three main modifications to our earlier model. First, we add an ex-ante investment stage during which foreign investors make their investment choices and after which their investments are sunk. This adds a new ex-ante inefficiency arising from distorted investment choices, which we allow the host government to address with up-front investment incentives. Second, the host government now applies an investment policy $\iota = \{FT, T\}$ (free trade or taking) to these sunk investments, where $\iota = T$ is a stand-in for a variety of investment policies (e.g., tax, regulatory, nationalization) that could amount to a "taking" broadly defined. This gives rise to an ex-post inefficiency similar to our earlier model, which the host government addresses through the BIT. And third, we model investment more explicitly than we modeled trade in the previous model, to be clear about the nature of the commitment problem as well as the various inefficiencies that can arise. Otherwise, the model essentially follows our earlier structure and we summarize the timing of events in Figure 3.

Just like real-world trade agreements, real-world investment agreements also tend to include vague language which is subject to interpretation. For example, Article 3 of the US Model BIT states:

> "Each Party shall accord to investors of the other Party treatment no less favorable than that it accords, in like circumstances, to its own investors with respect to the

establishment, acquisition, expansion, management, conduct, operation, and sale or other disposition of investments in its territory."

Here, what constitutes "like circumstances" is clearly a matter of interpretation, as might be the phrase "treatment no less favorable." Alternatively, we could think of the vague contract as stating the following (roughly corresponding to provisions in Article 6 of the US Model BIT):

"Foreign investments may be expropriated for a public purpose, provided that adequate and effective compensation is promptly paid."

Here, what constitutes "a public purpose," and what suffices for "adequate and effective compensation" paid "promptly," are matters of interpretation.

Model Preliminaries To fix ideas, we consider a specific FDI opportunity in the host country that, to exploit, requires a sunk capital investment by a risk-neutral foreign investor. To abstract from market access issues, we assume that the host country is small in world capital markets, so that it faces an infinitely elastic ex-ante supply of foreign capital at the world rate of return r^*. A single foreign investors makes a sunk investment I^*, which is converted one-for-one into output through the production function $Q = I^*$. We assume that the demand $D(P)$ for this output is elastic, $\frac{D'(P)P}{Q} < -1$, so that the investor always finds it optimal to sell all output ex-post. We denote the resulting market clearing price by $\tilde{P} = D^{-1}(I^*) \equiv \tilde{P}(I^*)$ and note that $\tilde{P}(I^*)$ is decreasing in I^* for $\tilde{P} > 0$.

Domestic consumer surplus, conditional on a level of FDI I^*, is then given by $CS(I^*) = \int_{\tilde{P}(I^*)}^{\infty} D(P)\,dP$ and is increasing in I^* with $CS'(I^*) = -\tilde{P}'(I^*)\,I^* > 0$. Similarly, the ex-post (conditional on sunk investment I^*) foreign operating profits or producer surplus is given by $PS(I^*) = \int_0^{\tilde{P}(I^*)} I^* dP = \tilde{P}(I^*)\,I^*$ and is increasing in I^* with $PS'(I^*) = \tilde{P}'(I^*)\,I^* + \tilde{P}(I^*) > 0$ given our assumption that demand is elastic. As a result, the sum of consumer and producer surplus in the market is also increasing in I^* with $CS'(I^*) + PS'(I^*) = \tilde{P}(I^*) > 0$.

Production or consumption of this output may generate a negative (local) externality that, while ignored by investors or individual consumers, could be large enough to turn the social value of the output from this investment negative. As in the model of the previous section, we assume that $s \equiv (s_1, s_2, \ldots, s_N)$ is a vector of state variables with each s_i corresponding to a binary event: here the realization of these state variables determines the magnitude of the

utility cost of the negative externality from the investment, which we denote by $e(I^*, s)$. For simplicity, we assume that $e(I^*, s) = e(s) I^*$ so that the utility cost is proportional to the size of the investment I^* with $e(s) \geq 0$ for all s.

We can partition the states of the world into those states where the externality is large enough to turn the ex-post social value of the investment negative, $PS(I^*)+CS(I^*)-e(I^*, s) < 0$, and those states where the ex-post social value of the investment remains positive, $PS(I^*) + CS(I^*) - e(I^*, s) \geq 0$. For simplicity we assume that in any state for which a negative externality is present (i.e., any s for which $e(s) > 0$), the externality $e(I^*, s)$ is large enough to ensure $PS(I^*)+CS(I^*)-e(I^*, s) < 0$ for any positive investment level I^*.[18] Hence, the states of the world may be partitioned into those where a negative consumption externality turns the ex-post social value of the investment negative, which we denote by $s \in \sigma^T$, and those where there is no externality ($e(s) = 0$) and the ex-post social value of the investment is positive and given by $PS(I^*) + CS(I^*)$, which we denote by $s \in \sigma^{FT}$.

Conditional on a given level of FDI, the investor's ex-post payoff is determined by the operating profits π that it collects; and these operating profits depend only on whether or not there is a "taking." If the investor is subject to a taking, the investor earns operating profits $\pi(I^*, T) = 0$ from the investment (regardless of the state of the world) implying a return on FDI of zero in that case. If the investor is not subject to a taking, the investor collects the market-clearing price $\tilde{P}(I^*)$ for the output from the investment (regardless of the state of the world) and therefore earns operating profits $\pi(I^*, FT) = \tilde{P}(I^*) I^* = PS(I^*)$ implying a return on FDI of $\tilde{P}(I^*)$ in that case.

By contrast, conditional on a given level of FDI, the domestic government's ex-post payoff depends both on the state of the world and on whether or not there is a taking. For $s \in \sigma^T$, the domestic government receives $\omega(I^*, T, s) = 0$ if there is a taking and the output from the investment is destroyed, and if there is not a taking the domestic government suffers the negative payoff $\omega(I^*, FT, s) = CS(I^*) - e(I^*, s) < 0$. For $s \in \sigma^{FT}$, the domestic government receives $\omega(I^*, T, s) = CS(I^*)+\kappa PS(I^*)$ if there is a taking and receives $\omega(I^*, FT, s) = CS(I^*)$ if there is not a taking, where $\kappa \in (0, 1)$ parameterizes the degree of ex-post inefficiency associated with the host-government's taking in σ^{FT}.[19] In particular, if κ were equal to one then the

[18] This amounts to an assumption that $e(s) > \tilde{P}(0)$ whenever $e(s) > 0$, where $\tilde{P}(0)$ is the "choke" price at which demand drops to zero. This simplifies the ensuing analysis but is not necessary for any of the results we emphasize.

[19] There is no role for the parameter κ when a taking occurs in σ^T, because by definition such takings are

taking would amount to a zero-sum transfer of surplus from the foreign investor to the Host government and would have no ex-post efficiency consequences (though, of course, possibly still ex-ante efficiency consequences for the level of investment), whereas when $\kappa \in (0,1)$ the host government gains less from the taking than the foreign investor loses and the taking therefore amounts to a negative-sum transfer between the parties. As we will show, the role of κ is central to the nature of the problem facing the host government and to the potential role of a BIT in helping to provide a solution to that problem.

This stylized description of a "taking" can be given a broad interpretation. The essential element is that states of the world exist in which the government action is socially efficient, and states of the world exist in which the government action is inefficient. For example, nationalization of an investor's property may transfer it to a higher valued use, or it may lead to a lower valued use. New regulations applicable to an investment may impose costs that are less than the social benefits, or may impose costs that exceed the social benefits. And so on.

We can define the ex-post (conditional on investment) gain that the host government enjoys from a taking in state s as $\gamma_G(I^*, s) \equiv \omega(I^*, T, s) - \omega(I^*, FT, s)$, and the lost rents suffered by the foreign investor in a taking as $\gamma^*_{I^*}(I^*) \equiv -PS(I^*)$. The joint ex-post gain from a taking for the host government and foreign investor is then given by $\Gamma(I^*, s) \equiv \gamma_G(I^*, s) + \gamma^*_{I^*}(I^*)$, and we have $\Gamma(I^*, s) = -[PS(I^*) + CS(I^*) - e(I^*, s)] > 0$ for $s \in \sigma^T$ and $\Gamma(I^*, s) = -(1-\kappa) PS(I^*) < 0$ for $s \in \sigma^{FT}$.

Hence, in states of the world $s \in \sigma^T$ the policy that maximizes the joint ex-post surplus for the host government and foreign investor, which we refer to as the "first best" policy, is complete expropriation (a "taking") and destruction of the output from the investment (T); and in states of the world $s \in \sigma^{FT}$ the first best policy is no expropriation, amounting to a government policy that allows the sale of the output from the investment to proceed unhindered (FT). Formally, and denoting by ι_{FB} the first-best policies taking I^* as given, we have $\iota_{FB} = FT$ for $s \in \sigma^{FT}$ and $\iota_{FB} = T$ for $s \in \sigma^T$.

Limited Commitment If the host government did not have access to any commitment technology, it would choose $\iota = T$ in all states of the world, expropriating the foreign investment in good states (σ^{FT}) and bad (σ^T), owing to the sunk nature of the FDI at the time that the host government makes its taking decision. This noncooperative policy choice would correspond to

ex-post efficient.

the first best for $s \in \sigma^T$, but it would differ from the first best for $s \in \sigma^{FT}$. Moreover, anticipating this ex-post treatment and hence a zero return on FDI, no foreign investment would be forthcoming, and the host government's welfare would be driven to zero in this market.

We assume that this stark commitment problem is mitigated by domestic institutions which limit the host government's ability to expropriate foreign investors. We have in mind domestic institutions that already protect the property rights of foreign investors, such as the domestic property law itself and the courts that enforce it. We capture this in a reduced-form fashion by assuming that the host government is forced to implement the first-best policies with probability $\bar{p}$ and can act at its own discretion with probability $1-\bar{p}$. We think of $\bar{p}$ as a parameter that varies across countries capturing differences in institutional quality, with $\bar{p} < 1$ signifying a lack of full commitment. Denoting by ι_C the policies implemented under this regime of limited commitment, we thus have $\iota_C = FT$ with probability $\bar{p}$ and $\iota_C = T$ with probability $1-\bar{p}$ for states $s \in \sigma^{FT}$ and $\iota_C = T$ in states $s \in \sigma^T$. Notice that those policies correspond to the first-best policies for $\bar{p} = 1$ and to the no-commitment policies for $\bar{p} = 0$.

Under limited commitment and absent any further policy interventions, the foreign investor therefore receives an expected return of $E_s\left[\rho_C^*\left(I^*, s\right)\right] = p^{FT}\bar{p}\tilde{P}\left(I^*\right)$, since the host government expropriates the foreign investor unless a state $s \in \sigma^{FT}$ realizes (which happens with probability $p^{FT} \equiv \sum_{s\in\sigma^{FT}} p\left(s\right)$) and the host government is constrained by domestic institutions to implement the first-best policy (which happens with probability $\bar{p}$). The equilibrium investment I_C^* is then implicitly defined by

$$p^{FT}\bar{p}\tilde{P}\left(I_C^*\right) = r^*, \tag{3.5}$$

since the foreign investor adjusts its investment to make the expected return equal its outside option r^*. This implies that the host government achieves an expected welfare of

$$E_s\left[\omega\left(I_C^*, \iota_C, s\right)\right] = p^{FT}\left[CS\left(I_C^*\right) + \left(1-\bar{p}\right)\kappa PS\left(I_C^*\right)\right], \tag{3.6}$$

since it shuts down production in states $s \in \sigma^T$, always enjoys the ex-post consumer surplus in states $s \in \sigma^{FT}$, and further seizes a fraction κ of the ex-post profits $PS\left(I_C^*\right)$ in states $s \in \sigma^{FT}$ if its hands are not tied by domestic institutions (which occurs with probability $1-\bar{p}$).

It is easy to show that $\frac{\partial E_s\left[\omega\left(I_C^*, \iota_C, s\right)\right]}{\partial \bar{p}} > 0$ by differentiating (3.5) and (3.6), which confirms that the host government indeed has a commitment problem vis-a-vis foreign investors. Intuitively, the foreign investor gets an expected return of r^* regardless of the host government's investment policy so that the host government ultimately pays the price for any deviation from

the first-best. A corollary of this is that the first-best policies also maximize the host government's expected welfare and not just the ex-post surplus of the host government and the foreign investor as established above.

The commitment problem that arises with $\bar{p} < 1$ gives rise to the possibility that, by controlling its ex-post incentive to expropriate and inducing more foreign investment into the domestic market, the host government could do better under a BIT. But it should also be clear that, in light of our assumption that the host country is small in world markets and therefore cannot impact the world interest rate r^*, the foreign investor stands to gain nothing from the BIT with the host country, as it can expect to earn the world interest rate on its investments wherever it invests.

Put differently, according to our model there is no inefficiency that can be traced to a *government-to-government international policy externality* in the absence of a BIT in the way there is in the absence of a trade agreement under our analysis of the previous subsection: instead, the commitment problem at the heart of investment agreements is a *government-to-investor policy commitment problem* that creates a host-country inefficiency, the costs of which are borne entirely by the host government. What a BIT can do is serve as a way for the host government to make policy commitments to foreign investors ex ante, and thereby address this host-country inefficiency. Hence, we can evaluate the BIT based on how close it comes to achieving $E_s\left[\omega\left(I^*_{FB}, \iota_{FB}, s\right)\right]$, the level of the host government's ex-ante expected welfare under the first-best policies ι_{FB}.

Investment Incentives As a preliminary step, we first ask whether the host government can correct this domestic inefficiency with a simpler policy response, namely, by offering an up-front investment incentive to the foreign investor to compensate it for the ex-post treatment to follow. Such a policy response is potentially appealing, because the host government can induce any desired level of foreign investment I^* by offering an up-front payment of $\left\{r^* - E_s\left[\rho_C\left(I^*, s\right)\right]\right\} I^*$ to foreign investors that is conditional on investing I^*. Since foreign investors expect a return on their investment of $E_s\left[\rho_C\left(I^*, s\right)\right]$ given the subsequent policy choices ι_C, the additional return $r^* - E_s\left[\rho_C\left(I^*, s\right)\right]$ makes the overall return from investing in the host country exactly equal to their outside option r^*.

Under such a program of up-front investment incentives, the host government's welfare for

any level of I^* would be equal to

$$\begin{aligned} E_s\left[\omega\left(I^*, \iota_C, s\right)\right] &= p^{FT}\left[CS\left(I^*\right) + \left(1-\bar{p}\right)\kappa PS\left(I^*\right)\right] - \left\{r^* - E_s\left[\rho_C\left(I^*, s\right)\right]\right\} I^* \\ &= p^{FT}\left[CS\left(I^*\right) + PS\left(I^*\right) - \left(1-\bar{p}\right)\left(1-\kappa\right) PS\left(I^*\right)\right] - r^* I^*. \end{aligned}$$

With probability p^{FT}, the foreign facility is not shut down once the state of the world is realized thus generating consumer surplus $CS\left(I^*\right)$. With additional probability $1-\bar{p}$, the host government is allowed to expropriate the foreign production facility which then further generates producer surplus of $\kappa PS\left(I^*\right)$. The host government's welfare consists of this expected surplus minus the costs of the up-front investment incentive scheme $\left\{r^* - E_s\left[\rho_C\left(I^*, s\right)\right]\right\} I^*$.

To determine the optimal level of investment $\hat{I}^*$ to induce under this program of up-front investment incentives, the host government solves $\max_{I^*} E_s\left[\omega\left(I^*, \iota_C, s\right)\right]$. This yields the first-order condition

$$\tilde{P}\left(\hat{I}^*\right) = \frac{r^*}{p^{FT}} + \left(1-\bar{p}\right)\left(1-\kappa\right)\frac{\partial PS\left(\hat{I}^*\right)}{\partial I^*}, \tag{3.7}$$

which implicitly defines $\hat{I}^*$. It is easy to verify using equations (**??**), (3.5), and (3.7) that the resulting level of investment $\hat{I}^*$ is smaller than the first-best level of investment I^*_{FB} but larger that the limited-commitment investment level I^*_C as long as $\left(1-\bar{p}\right)\left(1-\kappa\right) > 0$ since $\frac{\partial \tilde{P}\left(\hat{I}^*\right)}{\partial I^*} < 0$ and $\frac{\partial PS\left(\hat{I}^*\right)}{\partial I^*} > 0$.

Finally, with this optimal level of FDI secured by the appropriate up-front investment incentives program, the welfare enjoyed by the host government would then be given by

$$\begin{aligned} E_s\left[\omega\left(\hat{I}^*, \iota_C, s\right)\right] &= p^{FT}\left[CS\left(\hat{I}^*\right) + PS\left(\hat{I}^*\right) - \left(1-\bar{p}\right)\left(1-\kappa\right) PS\left(\hat{I}^*\right)\right] - r^*\hat{I}^* \quad (3.8) \\ &= p^{FT}\left[CS\left(\hat{I}^*\right) + \left(1-\bar{p}\right)\left(1-\kappa\right)\frac{\partial \tilde{P}\left(\hat{I}^*\right)}{\partial I^*}\left(\hat{I}^*\right)^2\right]. \end{aligned}$$

Equations (3.6) and (3.8) immediately imply that the up-front investment program improves on the limited commitment scenario as long as $\left(1-\bar{p}\right)\left(1-\kappa\right) > 0$ since the government then optimally chooses $\hat{I}^* > I^*_C$. It is also easy to see from equations (**??**) and (3.8) that the up-front investment program cannot achieve the first-best outcome since:

$$\begin{aligned} E_s\left[\omega\left(\hat{I}^*, \iota_C, s\right)\right] &\leq p^{FT} CS\left(\hat{I}^*\right) \\ &= E_s\left[\omega\left(\hat{I}^*, \iota_{FB}, s\right)\right] \\ &\leq E_s\left[\omega\left(I^*_{FB}, \iota_{FB}, s\right)\right]. \end{aligned}$$

We summarize with:

Lemma 2. *An up-front investment incentive program can help solve the host-government commitment problem with regard to foreign investors, but it cannot by itself achieve the first best.*

According to Lemma 2, up-front investment incentive programs fall short of being able to fully solve the host-government's commitment problem.[20] Nevertheless, up-front investment incentives are a relatively straightforward tool for attracting foreign investors that is commonly used in practice (see, for example, the analysis of tax holidays offered for this purpose in Bond, 1986). The question is then whether the host government can improve upon this relatively simple stand-alone policy response by going further and introducing as well a BIT. This is the question to which we now turn.

Investment Agreements As for what kind of a BIT is feasible, we make the same assumptions as we did with our analysis of trade agreements. Therefore, once a level of FDI I^* has been sunk, the model works just like the model of the previous section, with the host government now playing the role of the importer government from the previous section and the foreign investor/foreign government now playing the role of the foreign exporter/foreign government of the previous section. More specifically, beginning with a sunk level of I^* and proceeding as before, we can derive the equilibrium policy choices and filing behavior conditional on the complainant's cost of filing and payoff from winning in court. And in combination with an up-front investment incentive program that ensures for a given level of FDI I^* an expected return on investment r^* in light of the ensuing treatment of investment under the BIT, the host government can then optimize the up-front investment incentive to induce the investment level that maximizes expected host-government welfare under the BIT. We may then evaluate how the allocation of standing impacts the expected welfare of the host government by considering how this choice impacts the complainant's cost of filing and payoff from winning in court.

Accordingly, we continue to denote by $c_f^*(\cdot)$ the cost incurred by foreign complainant $f \in \{G^*, I^*\}$ whenever it invokes the DSB, but we now allow this cost to be a function of the level of

[20] In fact, our model overstates the likely effectiveness of up-front investment incentives in reality to solve the host government's ex ante commitment problem, because we abstract from incentive issues on the side of the investor (e.g., to deliver worthless investment once the up-front payment has been received). As we will show, this serves a useful pedagogical purpose, in that it allows for a clean targeting of the up-front investment incentive to the ex ante investment inefficiency and permits targeting the BIT design to addressing the ex-post (conditional on investment) inefficiency.

investment I^*, and in particular we assume that this cost rises in proportion to the magnitude of the producer surplus (operating profits) that is at stake in the taking. Formally, we assume that the cost incurred by foreign complainant f if it invokes the DSB in state s is given by $c_f^*(I^*, s) \equiv [c_f^* \epsilon^*(s)] PS(I^*)$; and denoting by $|\gamma_f^*(I^*)|$ the payoff to foreign complainant f if the DSB rules for FT, we assume that $|\gamma_f^*(I^*)| \equiv \gamma_f^* PS(I^*)$ where γ_f^* is a positive constant that parameterizes the complainant's payoff at stake in a taking. We make the analogous assumption for the host (defendant) government: if the host government is taken to court in state s, it incurs a cost $c(I^*, s) \equiv [c\epsilon(s)]\gamma_G(I^*, s)$ to defend the taking and enjoys the payoff $\gamma_G(I^*, s)$ if the DSB rules for T. After characterizing the equilibrium behavior conditional on $c_f^*(I^*, s)$ and $|\gamma_f^*(I^*)|$, the choice of standing will then amount to selecting the $f \in \{G^*, I^*\}$ that delivers the highest level of expected welfare for the host government, with $f = G^*$ corresponding to SSDS and $f = I^*$ corresponding to ISDS.

As before, we assume that the realized state s is observed by the governments and by the DSB, and that Γ is observed by the governments but not by the DSB (and we are assuming implicitly that the DSB cannot observe what the host government does with the production facility if it expropriates it, i.e., whether or not the production facility is shut down). As in the previous subsection, we will think of the DSB as issuing a policy ruling, in the present context denoted by ι^{DSB} and corresponding either to FT or T, to maximize the expected joint payoff of the host government and foreign investors given its noisy signal of Γ.

Under the interpretation that T represents a "regulatory taking," the DSB ruling could be seen in a richer model (as we discussed in the previous subsection, see note **??**) as corresponding to a determination of whether the regulation complies with some explicit (but vaguely worded) commitment included in the contract, such as national treatment or the MFN clause, which itself can be interpreted as an attribute of internationally efficient policy intervention. Alternatively, under the interpretation that T represents an explicit expropriation, the decision to expropriate could be left in the hands of the host government subsequent to the DSB ruling under the interpretation that the DSB rules on a level of compensation to be paid by the host government to the foreign investors in the event of expropriation, with the ruling FT then corresponding to a level of compensation sufficiently high to prevent the host government from following through with the expropriation and the ruling T corresponding to a level of compensation (which could be set arbitrarily to zero) under which the host government would go through

with expropriation.[21]

Consider first the foreign complainant's filing behavior. The complainant files a complaint if and only if $\iota = T$ and the expected benefit to the complainant of filing exceeds the complainant's cost of filing, that is

$$\Pr(\text{DSB ruling is } FT \mid s) \times |\gamma_f^*(I^*)| > c_f^*(I^*, s). \tag{3.9}$$

Condition (3.9) is the "filing" condition for the agent with standing in the foreign country to invoke the DSB in response to a policy choice by the host government of $\iota = T$.

Next consider the host government's policy choice, keeping in mind now that the host government is constrained to implement the first-best policies with probability $\bar{p}$. If the host government is constrained, it chooses $\iota = \iota_{FB}$. Otherwise, it chooses $\iota = T$ if either (3.9) fails – because then the host government can set $\iota = T$ without triggering a dispute – or if (3.9) holds and the expected benefit to the host government from a taking exceeds the cost to the host government of a dispute:

$$\Pr(\text{DSB ruling is } T \mid s) \times \gamma_G(I^*, s) > c(I^*, s). \tag{3.10}$$

We can now derive the equilibrium actions, conditional on investment level I^*, for each state s. For simplicity and as before, in what follows we assume that the states where the vague contract is unambiguous are measure zero, so we can focus only on states where the court if invoked must interpret the contract. Defining the thresholds $\bar{\mu}_1^{FT}(s) \equiv c\epsilon(s)$, $\bar{\mu}_{2,f}^{FT}(s) \equiv 1 - \frac{c_f^*}{\gamma_f^*}\epsilon^*(s)$, $\bar{\mu}_{1,f}^{T}(s) \equiv \frac{c_f^*}{\gamma_f^*}\epsilon^*(s)$, and $\bar{\mu}_2^T(s) \equiv 1 - c\epsilon(s)$, and noting as before that $\bar{\mu}_1^{FT}(s) < \bar{\mu}_{2,f}^{FT}(s)$ and $\bar{\mu}_{1,f}^{T}(s) < \bar{\mu}_2^T(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (3.9) and (3.10) immediately imply the following result:

[21] More specifically, in the case of explicit expropriation there would in practice typically be no question that this expropriation has occurred, and the main legal question before the court is simply to determine the level of damages. To map our model over to this case, and in analogy with our simplification of two policies T and FT, we assume that there are two possible levels of damages associated with expropriation, $High$ and Low, and σ^T then corresponds to states of the world where damages are Low and expropriation is efficient, while σ^{FT} corresponds to states of the world where damages are $High$ and expropriation is inefficient. And we assume that the host country would choose to expropriate in every state of the world if it only had to pay Low damages but would never choose to expropriate in any state of the world if it had to pay $High$ damages. With these assumptions, if the host country expropriates and the foreign investor invokes the court, then if the court rules for Low damages the host country will pay the Low damages and maintain its decision to expropriate (the analogue of a ruling of T, which is efficient if the state is in σ^T but inefficient if the state is in σ^{FT}), while if the court rules for $High$ damages the host country will reverse its decision to expropriate (give back the property to the foreign investors) to avoid paying the high damages (the analogue of a ruling of FT, which is efficient if the state is in σ^{FT} but inefficient if the state is in σ^T).

Lemma 3. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\bar{\mu}_1^{FT}(s) < \bar{\mu}_{2,f}^{FT}(s)$ and $\bar{\mu}_{1,f}^{T}(s) < \bar{\mu}_2^{T}(s)$, equilibrium actions can be characterized as follows:*

1. *In states $s \in \sigma^{FT}$:*

    1. *If the host government is constrained: We have $\iota = FT$ and no dispute.*
    2. *If the host government is unconstrained: If DSB quality is high in the sense that $qk(s) < \bar{\mu}_1^{FT}(s)$, we have $\iota = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\bar{\mu}_1^{FT}(s), \bar{\mu}_{2,f}^{FT}(s)\right]$, we have $\iota = T$ and a dispute; if DSB quality is low in the sense that $qk(s) > \bar{\mu}_{2,f}^{FT}(s)$, we have $\iota = T$ and no dispute.*

2. *In states $s \in \sigma^{T}$:*

    1. *If the host government is constrained: We have $\iota = T$, no dispute if $qk(s) < \bar{\mu}_{1,f}^{T}(s)$, and a dispute if $qk(s) > \bar{\mu}_{1,f}^{T}(s)$.*
    2. *If the host government is unconstrained: If DSB quality is high in the sense that $qk(s) < \bar{\mu}_{1,f}^{T}(s)$, we have $\iota = T$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\bar{\mu}_{1,f}^{T}(s), \bar{\mu}_2^{T}(s)\right]$, we have $\iota = T$ and a dispute; if DSB quality is low in the sense that $qk(s) > \bar{\mu}_2^{T}(s)$, we have $\iota = FT$ and no dispute.*

Turning to the ex-ante stage where the level of I^* is determined, recall that if the investor is not subject to a taking, the investor collects the market-clearing price $\tilde{P}(I^*)$ for the output from the investment (regardless of the state of the world) and therefore earns operating revenue $\pi(I^*, FT) = PS(I^*)$ implying a return on FDI of $\tilde{P}(I^*)$ in that case, whereas if the investor is subject to a taking the investor gets nothing. We will also assume that the investor always pays the filing cost $c_f^*(I^*, s)$, either directly in the case of ISDS (i.e., when $f = I^*$) or indirectly via the lobbying payments necessary to induce the foreign government to file in the case of SSDS (i.e., when $f = G^*$). Using this and the ex-post treatment of FDI implied by the equilibrium outcomes above, if the investor were to anticipate this treatment then this would imply an expected return on a level of FDI I^* from expected ex-post operating profits, which we denote by $E_s\left[\rho_f^*(I^*, s)\right]$. Defining the sets $\bar{\sigma}_1^{FT} \equiv \left\{s \in \bar{\sigma}^{FT} \mid qk(s) < \bar{\mu}_1^{FT}(s)\right\}$, $\bar{\sigma}_{2,f}^{FT} \equiv \left\{s \in \bar{\sigma}^{FT} \mid qk(s) \in \left[\bar{\mu}_1^{FT}(s), \bar{\mu}_{2,f}^{FT}(s)\right]\right\}$, and $\bar{\sigma}_{3,f}^{FT} \equiv \left\{s \in \bar{\sigma}^{FT} \mid qk(s) > \bar{\mu}_{2,f}^{FT}(s)\right\}$, as well as $\bar{\sigma}_{1,f}^{T} \equiv \left\{s \in \bar{\sigma}^{T} \mid qk(s) < \bar{\mu}_{1,f}^{T}(s)\right\}$, $\bar{\sigma}_{2,f}^{T} \equiv \left\{s \in \bar{\sigma}^{T} \mid qk(s) \in \left[\bar{\mu}_{1,f}^{T}(s), \bar{\mu}_2^{T}(s)\right]\right\}$, and,

$\bar{\sigma}_3^T \equiv \left\{ s \in \bar{\sigma}^T \mid qk(s) > \bar{\mu}_2^T(s) \right\}$, we can write:

$$\begin{aligned}
E_s\left[\rho_f^*(I^*,s)\right] &= (1-\bar{p}) \sum_{s \in \left\{\bar{\sigma}_1^{FT} \cup \bar{\sigma}_3^T\right\}} p(s)\tilde{P}(I^*) \\
&+ (1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,f}^{FT}} p(s) \left\{ [1-qk(s)]\tilde{P}(I^*) - \frac{c_f^*(I^*,s)}{I^*} \right\} \\
&+ (1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,f}^{T}} p(s) \left[qk(s)\tilde{P}(I^*) - \frac{c_f^*(I^*,s)}{I^*} \right] \\
&+ \bar{p}p^{FT}\tilde{P}(I^*) \\
&+ \bar{p} \sum_{s \in \left\{\bar{\sigma}_{2,f}^T, \bar{\sigma}_3^T\right\}} p(s) \left[qk(s)\tilde{P}(I^*) - \frac{c_f^*(I^*,s)}{I^*} \right].
\end{aligned}$$

Finally, denoting by $\omega_f(I^*,s)$ the host government payoff in state s conditional on a level of investment I^* in the presence of a BIT with standing choice $f \in \{I^*, G^*\}$, and recalling that the investor must expect to receive a return of r^* once both the up-front payments and the expected ex-post operating profits are accounted for, we can use the equilibrium actions as characterized above to express the expected level of this payoff:

$$\begin{aligned}
E_s[\omega_f(I^*,s)] &= (1-\bar{p}) \sum_{s \in \bar{\sigma}_1^{FT}} p(s) CS(I^*) \\
&+ (1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,f}^{FT}} p(s)\{CS(I^*) + qk(s)[\kappa PS(I^*)] - c(I^*,s)\} \\
&+ (1-\bar{p}) \sum_{s \in \bar{\sigma}_{3,f}^{FT}} p(s)[CS(I^*) + \kappa PS(I^*)] \\
&+ (1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,f}^{T}} p(s)\{qk(s)[CS(I^*) - e(I^*,s)] - c(I^*,s)\} \\
&+ (1-\bar{p}) \sum_{s \in \bar{\sigma}_{3}^{T}} p(s)[CS(I^*) - e(I^*,s)] + \bar{p}p^{FT} CS(I^*) \\
&+ \bar{p} \sum_{s \in \{\bar{\sigma}_{2,f}^T, \bar{\sigma}_{3,f}^T\}} p(s)\{qk(s)[CS(I^*) - e(I^*,s)] - c(I^*,s)\} \\
&- \{r^* - E_s[\rho_f^*(I^*,s)]\} I^*.
\end{aligned}$$

Plugging the expression for $E_s\left[\rho_f^*\left(I^*,s\right)\right]$ into the above expression for $E_s[\omega_f(I^*,s)]$ yields

$$\begin{aligned}
E_s\left[\omega_f\left(I^*,s\right)\right] &= (1-\bar{p})\sum_{s\in\bar{\sigma}_1^{FT}} p\left(s\right)\left[CS\left(I^*\right)+PS\left(I^*\right)\right] \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}_{2,f}^{FT}} p\left(s\right)\left\{CS\left(I^*\right)+\left[1-\left(1-\kappa\right)qk\left(s\right)\right]PS\left(I^*\right)-c_f^*\left(I^*,s\right)-c\left(I^*,s\right)\right\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}_{3,f}^{FT}} p\left(s\right)\left[CS\left(I^*\right)+\kappa PS\left(I^*\right)\right] \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}_{2,f}^{T}} p\left(s\right)\left\{qk\left(s\right)\left[CS\left(I^*\right)+PS\left(I^*\right)-e\left(I^*,s\right)\right]-c_f^*\left(I^*,s\right)-c\left(I^*,s\right)\right\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}_3^{T}} p\left(s\right)\left[CS\left(I^*\right)+PS\left(I^*\right)-e\left(I^*,s\right)\right] \\
&+\bar{p}p^{FT}\left[CS\left(I^*\right)+PS\left(I^*\right)\right] \\
&+\bar{p}\sum_{s\in\left\{\bar{\sigma}_{2,f}^{T},\bar{\sigma}_3^{T}\right\}} p\left(s\right)\left\{qk\left(s\right)\left[CS\left(I^*\right)+PS\left(I^*\right)-e\left(I^*,s\right)\right]-c\left(I^*,s\right)-c_f^*\left(I^*,s\right)\right\} \\
&-r^*I^*.
\end{aligned} \tag{3.11}$$

The interpretation of the expression for $E_s\left[\omega_f\left(I^*,s\right)\right]$ in (3.11) is intuitive, once it is understood that the host-government must pay the foreign investor the amount r^*I^* in equilibrium, as reflected in the last line of this expression; and with this paid, it is as if the host government then keeps for itself all of the ex-post net-of-litigation-cost surplus generated by I^* according to the equilibrium behavior in the presence of a BIT with standing choice $f\in\{I^*,G^*\}$ as characterized above. The first five lines of (3.11) record this ex-post surplus for the five sets of states where this surplus is non-zero for the case that the host government is unconstrained in its policymaking. And the last two lines add the surplus that is generated if the government is constrained to implement the first-best policies.

Using the expression for $E_s\left[\omega_f\left(I^*,s\right)\right]$ in (3.11), we can now solve for the optimal level of FDI in the presence of a BIT with standing choice $f\in\{I^*,G^*\}$, which we denote by $\bar{I}_f^*$, defined implicitly by $\frac{\partial E_s[\omega_f(\bar{I}_f^*,s)]}{\partial I^*}=0$. The associated first order condition can be manipulated to yield

the following implicit characterization of $\bar{I}_f^*$:

$$\begin{aligned}
\tilde{P}\left(\bar{I}_f^*\right) &= \frac{r^*}{p^{FT}} \\
&+ (1-\bar{p})(1-\kappa)\frac{p_{3,f}^{FT} + \sum_{s\in\bar{\sigma}_{2,f}^{FT}} p(s)\, qk(s)}{p^{FT}} \frac{\partial PS\left(\bar{I}_f^*\right)}{\partial I^*} \\
&+ (1-\bar{p}) \sum_{s\in\left\{\bar{\sigma}_{2,f}^{T}, \bar{\sigma}_{2}^{FT}\right\}} \frac{p(s)}{p^{FT}} \left[\frac{\partial c_f^*\left(\bar{I}_f^*, s\right)}{\partial I^*} + \frac{\partial c\left(\bar{I}_f^*, s\right)}{\partial I^*}\right] \\
&+ (1-\bar{p}) \frac{\sum_{s\in\bar{\sigma}_{2,f}^{T}} p(s)\, qk(s)}{p^{FT}} \left[e(s) - \tilde{P}\left(\bar{I}_f^*\right)\right] + \sum_{s\in\bar{\sigma}_{3}^{T}} \frac{p(s)}{p^{FT}} \left[e(s) - \tilde{P}\left(\bar{I}_f^*\right)\right] \\
&+ \bar{p} \sum_{s\in\left\{\bar{\sigma}_{2,f}^{T}, \bar{\sigma}_{3}^{T}\right\}} \frac{p(s)}{p^{FT}} \left[\frac{\partial c_f^*\left(\bar{I}_f^*, s\right)}{\partial I^*} + \frac{\partial c\left(\bar{I}_f^*, s\right)}{\partial I^*}\right] \\
&+ \bar{p} \sum_{s\in\left\{\bar{\sigma}_{2,f}^{T}, \bar{\sigma}_{3}^{T}\right\}} \frac{p(s)\, qk(s)}{p^{FT}} \left[e(s) - \tilde{P}\left(\bar{I}_f^*\right)\right].
\end{aligned} \tag{3.12}$$

We can see from the terms on the right-hand-side of (3.12) that there are a number of forces that determine the optimal level of investment $\bar{I}_f^*$ in the presence of a BIT with standing choice $f \in \{I^*, G^*\}$. The term in the first line simply reflects the first-best benchmark forces: if it were the only term on the right-hand-side of (3.12), then this expression would imply $\bar{I}_f^* = \hat{I}_{FB}^*$. The other terms capture various inefficiencies that make $\bar{I}_f^* < \hat{I}_{FB}^*$, first considering the case in which the host government is unconstrained and then turning to the case in which the host government is constrained. The second line reflects the inefficiencies associated with takings in σ^{FT}, similar to those in (3.7) but down-weighted by the reduced probability of a taking in σ^{FT} under the BIT. The third line reflects the impact of greater investment on the expected costs of litigation, both complainant and defendant, which the host government must ultimately bear. And the terms on the fourth line, which are both positive, reflect the expected welfare costs that arise when efficient takings are blocked by incorrect court rulings. The terms in the last two lines make clear that there would still be inefficiencies even in the special case $\bar{p} = 1$, since the foreign investor then still litigates whenever $s \in \left\{\bar{\sigma}_{2,f}^{T}, \bar{\sigma}_{3}^{T}\right\}$ which brings about litigation costs and prevents efficient takings in case of court mistakes.

Finally, plugging $\bar{I}_f^*$ into the expression for $E_s\left[\omega_f\left(I^*,s\right)\right]$ in (3.11) and simplifying yields

$$\begin{aligned} E_s\left[\omega_f\left(\bar{I}_f^*,s\right)\right] &= (1-\bar{p})\left[p^{FT}+p_3^T+\sum_{s\in\bar{\sigma}_{2,f}^T}p\left(s\right)qk\left(s\right)\right]CS\left(\bar{I}_f^*\right) \\ &+(1-\bar{p})\left[p_{3,f}^{FT}+\sum_{s\in\bar{\sigma}_{2,f}^{FT}}p\left(s\right)qk\left(s\right)\right](1-\kappa)\left(\bar{I}_f^*\right)^2\frac{\partial\tilde{P}\left(\bar{I}_f^*\right)}{\partial\bar{I}_f^*} \\ &-(1-\bar{p})\sum_{s\in\left\{\bar{\sigma}_{2,f}^{FT}\cup\bar{\sigma}_{2,f}^T\right\}}p\left(s\right)\left\{c_f^*\left(\bar{I}_f^*,s\right)+c\left(\bar{I}_f^*,s\right)-\bar{I}_f^*\left[\frac{\partial c_f^*\left(\bar{I}_f^*,s\right)}{\partial I^*}+\frac{\partial c\left(\bar{I}_f^*,s\right)}{\partial I^*}\right]\right\} \\ &+\bar{p}\left[p^{FT}+\sum_{s\in\left\{\bar{\sigma}_{2,f}^T,\bar{\sigma}_3^T\right\}}p\left(s\right)qk\left(s\right)\right]CS\left(I^*\right) \\ &-\bar{p}\sum_{s\in\left\{\bar{\sigma}_{2,f}^T,\bar{\sigma}_3^T\right\}}p\left(s\right)\left\{c\left(I^*,s\right)+c_f^*\left(I^*,s\right)-\bar{I}_f^*\left[\frac{\partial c_f^*\left(\bar{I}_f^*,s\right)}{\partial I^*}+\frac{\partial c\left(\bar{I}_f^*,s\right)}{\partial I^*}\right]\right\}. \end{aligned} \tag{3.13}$$

Notice that the BIT only improves upon the program of up-front investment incentives if court quality is sufficiently high and domestic institutions are sufficiently weak. If $qk\left(s\right)>\max\left\{\bar{\mu}_{2,f}^{FT}\left(s\right),\bar{\mu}_2^T\left(s\right)\right\}$ for all s, the presence of the low-quality court and the threat of litigation serves only to shield inefficient behavior, with takings never occurring when they should (in σ^T) and always occurring with impunity when they should not (in σ^{FT}). In this case we have using (3.8) and (3.11) that $E_s\left[\omega_f\left(\bar{I}_f^*,s\right)\right]\leq E_s\left[\omega\left(\hat{I}^*,\iota_C,s\right)\right]$. Also, if $\bar{p}=1$ so that $E_s\left[\omega\left(\hat{I}^*,\iota_C,s\right)\right]=E_s\left[\omega\left(I_{FB}^*,\iota_{FB},s\right)\right]$, the BIT necessarily makes things worse as long as $qk\left(s\right)>\bar{\mu}_1^T\left(s\right)$ for some s so that there is some litigation in equilibrium. On the other hand, if $qk\left(s\right)<\min\left\{\bar{\mu}_1^{FT}\left(s\right),\bar{\mu}_1^T\left(s\right)\right\}$ for all s the sets $\left\{\bar{\sigma}_{2,f}^{FT},\bar{\sigma}_{3,f}^{FT},\bar{\sigma}_{2,f}^T,\bar{\sigma}_3^T\right\}$ become empty and thus $E_s\left[\omega_f\left(\bar{I}_f^*,s\right)\right]=E_s\left[\omega\left(I_{FB}^*,\iota_{FB},s\right)\right]>E_s\left[\omega\left(\hat{I}^*,\iota_C,s\right)\right]$.[22]

We now have an answer to the question we raised in our discussion following Lemma 2, namely, whether the host government can improve upon offering a stand-alone up-front investment incentive program by going further and introducing a BIT. We summarize with:

[22] As we have noted (see note 15), we are ignoring issues associated with the endogeneity of contract choice and court mandate and highlighted in the analysis of Maggi and Staiger (2011). But the point we emphasize would survive endogenizing the institutional choice in this way: our point is simply that with low enough court quality, it would not be optimal to introduce a BIT-like institution where a court is endowed with the ability to settle investment disputes and where these disputes do occur along the equilibrium path.

Proposition 2. *The introduction of a BIT can lead to efficiency gains and benefit the host government relative to a stand-alone program of offering up-front investment incentives to foreign investors if and only if the quality of the court is sufficiently high and the quality of domestic institutions is sufficiently weak.*

There is no analogue to this result (i.e., the possibility of a purely unilateral intervention that could dominate an international agreement) for the case of trade agreements, a difference that reflects the different nature of the inefficiencies addressed across the two kinds of agreements, a fundamentally domestic inefficiency in the case of BITs and a fundamentally international inefficiency in the case of trade agreements. Henceforth we will assume that court quality is sufficiently high and the host-country commitment problem is sufficiently severe to allow the introduction of a BIT to improve upon up-front investment incentives by themselves, at least when standing is optimally allocated in the BIT.

SSDS versus ISDS We now evaluate the desirability of adopting SSDS versus ISDS. Tracking the choice in actual BITs, the relevant thought experiment would be to compare a BIT institution that includes both SSDS and ISDS to an institution that includes only SSDS but no ISDS. But again, as we did for our analysis of trade agreements, for simplicity we choose to proceed formally in this more parsimonious way, and then draw observations relevant to the more realistic possibility of ISDS as an addition to (rather than a replacement for) SSDS in a BIT.

As we have noted, the relative merits of SSDS versus ISDS can be evaluated once we specify the complainant's cost of filing and payoff from winning in court under each choice of standing, that is, when only the foreign government can be the complainant as under SSDS, and when only the foreign investor can be the complainant as under ISDS.

In analogy with Assumption 1 of the previous subsection, we assume that the cost of filing for the government is the same as the cost of filing for the foreign investor, namely

$$c^*_{G^*} = c^*_{I^*} \equiv c^* \qquad \text{(Assumption 1}')$$

implying

$$c^*_{G^*}(I^*, s) = c^*_{I^*}(I^*, s) = [c^*\epsilon^*(s)]PS(I^*) \equiv c^*(I^*, s) \text{ for all } s.$$

Hence, as with our analysis of standing in trade agreements in the previous subsection, any difference between a BIT with SSDS and a BIT with ISDS must arise due to differences across

the complainants' payoffs from winning in court. Our key assumption is that the foreign government is an *imperfect agent* for the foreign investor, in the following sense: given any FDI level I^*, the foreign government's loss from a taking is less than the producer surplus that would be lost by foreign investors, that is

$$\gamma^*_{G^*} < \gamma^*_{I^*} = 1 \qquad \text{(Assumption 2}')$$

implying

$$|\gamma^*_{G^*}(I^*)| = \gamma^*_{G^*} PS(I^*) < PS(I^*) = |\gamma^*_{I^*}(I^*)|.$$

Assumption 2′ is the analog of Assumption 2 that we introduced in the context of trade agreements. There we motivated this assumption by appealing both to the winners and losers that exist within a country when it faces trade protection from abroad, and to the "political filter" that is lost if standing is extended beyond the foreign government. Here we have in mind that the diminished payoff that would be enjoyed by the foreign government from a court win relative to that enjoyed by the foreign investor as implied by Assumption 2′ could reflect both political costs borne by the foreign government and the opportunity costs of diverting public funds and resources for this purpose.[23]

Consider now how expected host-government welfare would be impacted by switching from a BIT with SSDS to one with ISDS. Asking the analogous question in the previous subsection with regard to the joint surplus generated by a trade agreement under SSDS versus ESDS, Proposition 1 reported that an optimally designed trade agreement should include SSDS, but not ESDS. Are there conditions under which in the present context ISDS could be part of an optimally designed BIT?

To answer this question, we do not derive expressions for the expected efficiency loss under ISDS and under SSDS relative to the first-best outcome as we did for the analysis of trade agreements in the previous subsection, because the changes in optimal investment levels under each

[23] For example, if the series of tobacco plain-packaging disputes brought by Philip Morris under the ISDS provisions of various US BITs had instead been brought by the United States government, it seems plausible in light of the public controversy surrounding these disputes that the United States government would have enjoyed diminished gains from a win at court relative to the gains that would be enjoyed by Philip Morris (reflecting, for example, the political costs of having the name of the United States government associated with litigation aimed at weakening the health regulations of other countries and dealing with various constituencies in the United States on these issues). We note also that we have modeled these differences as entering through the complainant's payoff of a win in court. An alternative would be to capture these differences in the complainant's cost function of bringing a dispute, with the foreign government then bearing higher costs of bringing a dispute than the foreign investor. The results that we emphasize below concerning conditions under which ISDS is optimal for a BIT would only be strengthened under this alternative modeling approach, as in that case there would also be a direct dispute costs savings from adopting ISDS rather SSDS.

policy regime complicate such comparisons in the context of BITs. Instead we hold investment at the optimal level under a BIT with SSDS ($\bar{I}^*_{G^*}$) as characterized by (3.12), and we use (3.11) to calculate $E_s[\omega_{I^*}(\bar{I}^*_{G^*}, s)] - E_s[\omega_{G^*}(\bar{I}^*_{G^*}, s)]$, seeking conditions under which this difference is positive, which then provide sufficient conditions for $E_s[\omega_{I^*}(\bar{I}^*_{I^*}, s)] > E_s[\omega_{G^*}(\bar{I}^*_{G^*}, s)]$ and hence for ISDS to be part of an optimal BIT, given that $E_s[\omega_{I^*}(\bar{I}^*_{I^*}, s)] \geq E_s[\omega_{I^*}(\bar{I}^*_{G^*}, s)]$.

To express $E_s[\omega_{I^*}(\bar{I}^*_{G^*}, s)] - E_s[\omega_{G^*}(\bar{I}^*_{G^*}, s)]$, we proceed as above and first define the sets $\bar{\Omega}^{FT} \equiv \left\{ s \in \sigma^{FT} \mid qk(s) \in \left[\bar{\mu}^{FT}_{2,G^*}(s), \bar{\mu}^{FT}_{2,I^*}(s) \right] \right\}$ and $\bar{\Omega}^{T} \equiv \left\{ s \in \sigma^{T} \mid qk(s) \in \left[\bar{\mu}^{T}_{1,I^*}(s), \bar{\mu}^{T}_{1,G^*}(s) \right] \right\}$. As defined, the set $\bar{\Omega}^{FT}$ describes states in σ^{FT} where the host government would implement a taking with impunity under SSDS but would face litigation under ISDS (i.e., states that are in $\bar{\sigma}^{FT}_{3,G^*}$ under SSDS but switch to $\bar{\sigma}^{FT}_{2,I^*}$ under ISDS). Similarly, the set $\bar{\Omega}^{T}$ describes states in σ^{T} where the host government would implement a taking without court challenge under SSDS but would face litigation under ISDS (i.e., states that are in $\bar{\sigma}^{T}_{1,G^*}$ under SSDS but switch to $\bar{\sigma}^{T}_{2,I^*}$ under ISDS). With these new sets defined and using (3.11), we can now write $\Delta_{BIT} \equiv E_s[\omega_{I^*}(\bar{I}^*_{G^*}, s)] - E_s[\omega_{G^*}(\bar{I}^*_{G^*}, s)]$ as

$$\begin{aligned} \Delta_{BIT} &= (1-\bar{p}) \sum_{s \in \bar{\Omega}^{FT}} p(s) \left\{ (1 - qk(s))(1-\kappa) PS\left(\bar{I}^*_{G^*}\right) - c^*(\bar{I}^*_{G^*}, s) - c\left(\bar{I}^*_{G^*}, s\right) \right\} \qquad (3.14) \\ &\quad - \sum_{s \in \bar{\Omega}^{T}} p(s) \left\{ qk(s) \left[e\left(\bar{I}^*_{G^*}, s\right) - CS\left(\bar{I}^*_{G^*}\right) - PS\left(\bar{I}^*_{G^*}\right) \right] + c^*\left(\bar{I}^*_{G^*}, s\right) + c\left(\bar{I}^*_{G^*}, s\right) \right\}. \end{aligned}$$

The sign of the term on the first line of (3.14) is generally ambiguous and captures the welfare effects of additional litigation under ISDS in states $s \in \bar{\Omega}^{FT}$ when the host government is unconstrained. The upside of this additional litigation is that a share $1-\kappa$ of producer surplus gets saved if the court issues the correct ruling and prevents the host government from taking over production. The downside is that the additional litigation is costly for both countries. The overall sign of this term depends on the value of κ and $c\left(\bar{I}^*_{G^*}, s\right) = c\epsilon(s)\gamma_G(I^*, s)$: as $\kappa \to 0$ and $c \to 0$, this term becomes unambiguously positive since then $(1 - qk(s)) PS\left(\bar{I}^*_{G^*}\right) > c^*(\bar{I}^*_{G^*}, s)$ from the filing condition for $s \in \bar{\Omega}^{FT}$. The term on the second line is unambiguously negative and captures the welfare loss from additional litigation in states $s \in \bar{\Omega}^{T}$ brought about by court mistakes and litigation costs. Overall, we therefore have $\Delta_{BIT} > 0$ if (i) domestic institutions are sufficiently bad ($\bar{p}$ low), (ii) there are large efficiency losses from expropriation (κ small), (iii) the host government's costs from defending itself in court are small (c small), and (iii) states $s \in \bar{\Omega}^{T}$ are sufficiently unlikely (which is guaranteed if p^T is small).

Finally, as we noted at the outset of this subsection, we have proceeded with our formal

analysis by considering the choice of either ISDS or SSDS to be included in a BIT. In reality, the relevant design choice is more aptly described as whether to include ISDS in a BIT in addition to SSDS. Suppose we assume that when both are included, ISDS filing trumps SSDS filing whenever both are incentivized to file. Then our discussion above applies equally well to this design choice. We may now state:

Proposition 3. *Whether investors, in addition to their governments, should have standing to bring disputes in an optimally designed BIT depends on a number of subtle tradeoffs. But if the host government is highly inefficient in orchestrating takings for $s \in \sigma^{FT}$ and bears little cost of defending itself in court, and if domestic institutions are sufficiently weak and expropriation is socially efficient only in unusual circumstances, then it is optimal to give investors standing to bring disputes in a BIT. That is, if κ, c, $\bar{p}$ and p^T are sufficiently low, an optimally designed BIT should include both an SSDS and an ISDS.*

The findings reported in Proposition 3 are highly intuitive. If inefficient takings create large inefficiencies, then litigation to avert them can bring substantial gains. If litigation costs for defendants are lower, the costs of increased litigation under ISDS relative to SSDS are less burdensome. If domestic institutions are weak, they cannot be counted on to protect investors. And if "expropriation" is efficient only infrequently, the probability of inefficient litigation over desirable takings is diminished. Bear in mind that the quality of the court sits in the background of Proposition 3 as well – the benefits of a BIT, and of ISDS, increase with the accuracy of adjudication.[24]

We emphasize that we are focused here on sufficient conditions for ISDS to be desirable. Proposition 3 nevertheless indicates that the case for ISDS provisions in BITs is far from absolute, and whether the conditions for its optimal inclusion in BITs are met in practice may not be obvious.[25] We thus interpret Proposition 3 as providing valuable insight into why ISDS

[24] An interesting question is why the negative externality that the foreign government imposes on the host government when it files does not guarantee overly litigious behavior in the context of a BIT as it did in the context of a trade agreement, and thereby guarantee that giving standing to a more aggressive filer would diminish the value of a BIT as it did in the context of a trade agreement. The answer is that the optimal design of a trade agreement maximizes the ex-ante joint surplus of the two governments and hence the filing externality between them is directly relevant to the ex-ante joint payoff that the trade agreement is designed to maximize, while in the case of a BIT it is simply the host government's ex-ante payoff that is being maximized and so the implication of a negative externality experienced by the host government when the foreign government files is not directly relevant for the ex-ante payoff that the BIT is designed to maximize.

[25] In fact, holding investment fixed at the optimal level under a BIT with ISDS, analogous arguments to those leading up to Proposition 3 can be used to describe as well sufficient conditions for it to be optimal not to include ISDS in a BIT (i.e, roughly the opposite of those described in Proposition 3).

has become quite controversial.

We also underscore that while Proposition 3 describes conditions under which ISDS can be a valuable feature of a BIT, it stands in stark contract to Proposition 1 which rules out the use of ESDS in an optimal trade agreement. In this sense our results can help to explain the common policy of affording private standing in BITs but not trade agreements: at a fundamental level this difference reflects the different nature of the central problem addressed by each kind of agreement (market access versus credible commitment).

Standing for disputes over commitments to investors more generally We have analyzed commitment issues with respect to foreign investors within the context of BITs. Arguably, similar issues may arise in the context of trade agreements: indeed, Yarbrough and Yarbrough (1992) argue that a central role for trade agreements is to allow importer governments to make policy commitments to foreign exporters who must make sunk investments in order to export to their markets.[26] In principle, our analysis above (with the interpretation now excluding the possibility of explicit expropriation) could be applied directly to trade agreements wherever these agreements are designed to address such commitment issues, with an ESDS mechanism playing the role in trade agreements that is played by ISDS mechanisms in BITs.

However, it may be plausible to view such commitment issues as less important in the context of trade agreements than they are in the context of BITs, for the simple reason that the issue of sunk investments may be more important in the context of FDI than in the context of exporting. There are several reasons to think that this distinction may be important. First, and most obviously, there is a lack of any direct expropriation threat to the investments of exporters, in contrast to the case for FDI. But beyond this, in a multi-country world the investments made by exporters will commonly have alternative uses to produce exports to other markets – and to this extent therefore not be sunk – whereas FDI would continue to be largely sunk and therefore highly susceptible to hold-up by the host country in a multi-country world.

If one accepts this distinction, then it follows that the ex ante investment problem is more important in the context of BITs than it is in the context of trade agreements. And if setting up ISDS system involves some fixed-cost component so that it is not worth doing below some

[26] See also McLaren (1997) whose analysis of a trade agreement between a large and a small country turns this argument for trade agreements on its head.

minimal level of hold-up threat, then this distinction could account for the inclusion of ISDS provisions in BITs when the conditions of Proposition 3 are satisfied but no analogous inclusion of ESDS provisions in trade agreements to handle commitment issues. Likewise, if it is difficult to create an ESDS mechanism that limits private standing to cases where serious commitment problems arise, but denies it for market access disputes, this would provide a further rationale for the exclusion of ESDS from trade agreements.

4. Nature of the Remedy

We have assumed thus far that damage payments are not part of the court's ruling - when a case is filed and the court sides with the complainant, the defendant has no choice but to "cease and desist" whatever policy is found to be illegal. In reality, however, convicted violators of trade agreements and BITs have the alternative option to continue their violation and compensate the injured party through some form of damage payment. There is a fundamental difference in the nature of damage payments in trade agreements and BITs which is the source of much public controversy: while trade agreements allow the injured party to engage in reciprocal retaliation, BITs provide explicitly for cash payments.

In this section, we explore this difference in the nature of the remedy. To this end, we extend our baseline models from section 3 to allow for damage payments. Rather than simply assuming that the Home or Host government switches to $\tau = FT$ or $\iota = FT$ when convicted by the DSB, we now allow the government to maintain $\tau = P$ or $\iota = T$ but make damage payments. We compare two institutional setups: one in which the court allows reciprocal retaliation and another in which the court awards cash damages. The key trade-off featured by our extended model is that retaliation is less efficient but that cash damages are harder to assess. Which institution is optimal then depends on which force is stronger. In effect, we will argue that it makes sense that BITs provide for cash payments while trade agreements do not, provided that the cash value of the harm suffered by a foreign investor who is subject to a taking is sufficiently easy to quantify relative to the cash value of the harm suffered by a foreign government whose exporters face trade protection.

4.1. Trade Agreements

We first consider the choice of retaliation versus cash payments in the context of trade agreements. Building on the findings of the previous section, we assume that the foreign government

has standing in trade disputes. To make our points as clearly as possible, we adopt two further simplifying assumptions. First, when retaliation is the remedy (which we refer to as the V_R *institution*) we assume that retaliation is so inefficient and costly that Home always switches to $\tau = FT$ when convicted by the DSB to avoid retaliation - this means that in the V_R institution everything is exactly the same as in the V_{G^*} institution of our baseline model. Second, when cash payments are the remedy (which we refer to as the V_C *institution*) we assume that cash payments are perfectly efficient so that consumer surplus and producer surplus can be costlessly transferred internationally.

To capture the notion that the court struggles to accurately assess cash damages, we assume that the court-assessed damages are realizations of a random variable. Denoting the damages awarded to the injured foreign government by $d^*_{G^*}(s)$, we assume $\Pr\left[d^*_{G^*}(s) = |\gamma^*_{G^*}(s)|\right] = 1 - 2m(s)$, $\Pr\left[d^*_{G^*}(s) > |\gamma^*_{G^*}(s)|\right] = m(s)$, and $\Pr\left[d^*_{G^*}(s) < |\gamma^*_{G^*}(s)|\right] = m(s)$, with $m(s) \in [0, \frac{1}{2}]$. Hence, the court awards the correct damages with probability $1 - 2m(s)$ and overestimates or underestimates the damages with symmetric probabilities $m(s)$. Defining $\bar{d}^*_{G^*} \equiv E\left[d^*_{G^*}(s) \mid d^*_{G^*}(s) > |\gamma^*_{G^*}(s)|\right]$ and $\underline{d}^*_{G^*} \equiv E\left[d^*_{G^*}(s) \mid d^*_{G^*}(s) < |\gamma^*_{G^*}(s)|\right]$, we further assume that $\bar{d}^*_{G^*} > \gamma_G(s)$ and $\underline{d}^*_{G^*} < \gamma_G(s)$. So, if the court overestimates the damages, Home can expect to be assessed damages in the amount of $\bar{d}^*_{G^*}$, which would be higher than Home's valuation of the violation $\gamma_G(s)$. Conversely, if the court underestimates the damages, Home can expect to be assessed damages in the amount of $\underline{d}^*_{G^*}$ which would be lower than its valuation of the violation $\gamma_G(s)$. This ensures that damage assessment mistakes are consequential in the sense that they influence the choice of Home to cease and desist or pay damages.

Notice also that the accuracy with which the court assesses damages ($d^*_{G^*}(s)$), as parameterized by $m(s)$, is distinct from the accuracy of the court ruling (FT or P), as parameterized by $qk(s)$. Thus, for example, the court might be good at determining whether or not the imposition of protection was warranted in a particular state of the world (e.g., Does protection preserve more jobs in the Home country than it destroys in the Foreign country?) but bad at assessing the value of the harm done to the foreign government (e.g., What is the monetary value of a job?).

With equilibrium actions under the V_R institution exactly the same as in the V_{G^*} institution of our baseline model, it only remains to characterize the equilibrium actions under the V_C institution. Under the V_C institution, Foreign files a complaint if $\tau = P$ and

$$\Pr(\text{ruling is } FT \mid s)\left\{\Pr(\tau = FT)\left|\gamma^*_{G^*}(s)\right| + \Pr(\tau = P)\, E\left[d^*_{G^*}(s) \mid \tau = P\right]\right\} > c^*(s) \quad (4.1)$$

and Home chooses $\tau = P$ if either the above condition is violated or

$$\Pr(\text{ruling is } P \mid s)\,\gamma_G(s) + \Pr(\text{ruling is } FT \mid s)\Pr(\tau = P)\left\{\gamma_G(s) - E\left[d^*_{G^*}(s) \mid \tau = P\right]\right\} > c(s). \tag{4.2}$$

Here, $\Pr(\tau = FT)$ is the probability that Home chooses to cease and desist in the face of a court ruling of FT while $\Pr(\tau = P)$ is the probability that Home instead continues with $\tau = P$ and pays damages, so that $\Pr(\tau = FT) = 1 - \Pr(\tau = P)$. What is new relative to our earlier equations (3.1) and (3.2) is that we allow for $\Pr(\tau = P) > 0$. So, when Home and Foreign now decide on their actions, they have to account for the possibility that Home may choose to continue the violation and make damage payments to Foreign.

Based on these considerations, we can then derive the equilibrium actions for each state s. Defining the thresholds $\mu^{FT}_{1,C}(s) \equiv \frac{c(s)-m(s)\left[\gamma_G(s)-\underline{d}^*_{G^*}\right]}{\gamma_G(s)-m(s)\left[\gamma_G(s)-\underline{d}^*_{G^*}\right]}$, $\mu^{FT}_{2,C}(s) \equiv 1 - \frac{c^*(s)}{\left|\gamma^*_{G^*}(s)\right|-m(s)\left[\left|\gamma^*_{G^*}(s)\right|-\underline{d}^*_{G^*}\right]}$, $\mu^{P}_{1,C}(s) \equiv \frac{c^*(s)}{\left|\gamma^*_{G^*}(s)\right|-m(s)\left[\left|\gamma^*_{G^*}(s)\right|-\underline{d}^*_{G^*}\right]}$, and $\mu^{P}_{2,C}(s) \equiv \frac{\gamma_G(s)-c(s)}{\gamma_G(s)-[1-2m(s)]\left[\gamma_G(s)-\left|\gamma^*_{G^*}(s)\right|\right]-m(s)\left[\gamma_G(s)-\underline{d}^*_{G^*}\right]}$, and noting as before that $\mu^{FT}_{1,C}(s) < \mu^{FT}_{2,C}(s)$ and $\mu^{P}_{1,C}(s) < \mu^{P}_{2,C}(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (4.1) and (4.2) immediately imply the following result:

Lemma 4. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\mu^{FT}_{1,C}(s) < \mu^{FT}_{2,C}(s)$ and $\mu^{P}_{1,C}(s) < \mu^{P}_{2,C}(s)$, equilibrium actions under the V_C institution can be characterized as follows:*

1. *In states $s \in \sigma^{FT}$: If DSB quality is high in the sense that $qk(s) < \mu^{FT}_{1,C}(s)$, we have $\tau = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\mu^{FT}_{1,C}(s), \mu^{FT}_{2,C}(s)\right]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu^{FT}_{2,C}(s)$, we have $\tau = P$ and no dispute.*

2. *In states $s \in \sigma^{P}$: If DSB quality is high in the sense that $qk(s) < \mu^{P}_{1,C}(s)$, we have $\tau = P$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\mu^{P}_{1,C}(s), \mu^{P}_{2,C}(s)\right]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu^{P}_{2,C}(s)$, we have $\tau = FT$ and no dispute.*

Notice that the basic structure of the actions is the same as in the V_{G^*} institution of our baseline model (and hence the same as in the the V_R institution). The court induces the first-best action in states where it is highly accurate, shields the inefficient action in states where it

is highly inaccurate, and gets actively involved in states in between. Hence, damage payments are made in equilibrium only in states of intermediate court quality, though the option of making them affects all action thresholds. For future reference, we denote the different action sets by $\hat{\sigma}_{1,C}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) < \mu_{1,C}^{FT}(s)\right\}$, $\hat{\sigma}_{2,C}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) \in \left[\mu_{1,C}^{FT}(s), \mu_{2,C}^{FT}(s)\right]\right\}$, and $\hat{\sigma}_{3,C}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) > \mu_{2,C}^{FT}(s)\right\}$, as well as $\hat{\sigma}_{1,C}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) < \mu_{1,C}^{P}(s)\right\}$, $\hat{\sigma}_{2,C}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) \in \left[\mu_{1,C}^{P}(s), \mu_{2,C}^{P}(s)\right]\right\}$, and, $\hat{\sigma}_{3,C}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) > \mu_{2,C}^{P}(s)\right\}$.

We can now characterize the efficiency loss associated with the V_C institution, $L(V_C)$, relative to the first-best outcome of $\tau = FT$ if $s \in \sigma^{FT}$ and $\tau = P$ if $s \in \sigma^{P}$. There are two main changes relative to our baseline analysis. First, if the court correctly rules FT there is now still a probability $m(s)$ that it awards excessively low damages $\underline{d}_{G^*}^{*}$ in which case Home will continue to choose $\tau = P$, leading to an additional efficiency loss. Second, if the court incorrectly rules FT there is still a probability $1 - 2m(s) + m(s) = 1 - m(s)$ that it does not award excessively high damages in which case Home will continue to choose $\tau = P$, leading to an additional efficiency gain:

$$\begin{aligned} L(V_C) &= \sum_{s \in \hat{\sigma}_{2,C}^{FT} \cup \hat{\sigma}_{2,C}^{P}} p(s)\left[qk(s)\left|\Gamma(s)\right| + c(s) + c^{*}(s)\right] \\ &+ \sum_{s \in \hat{\sigma}_{3,C}^{FT} \cup \hat{\sigma}_{3,C}^{P}} p(s)\left|\Gamma(s)\right| \\ &+ \sum_{s \in \hat{\sigma}_{2,C}^{FT}} p(s)\left[1 - qk(s)\right] m(s)\left|\Gamma(s)\right| \\ &- \sum_{s \in \hat{\sigma}_{2,C}^{P}} p(s)\, qk(s)\left[1 - m(s)\right]\left|\Gamma(s)\right|. \end{aligned}$$

And finally, given our assumption that retaliation is so inefficient and costly that Home always switches to $\tau = FT$ when convicted by the DSB to avoid retaliation ($\Pr(\tau = P) = 0$), the efficiency loss associated with the V_R institution, $L(V_R)$, is exactly the same as in the V_{G^*} institution of our baseline model. To aid comparison, we use the notation $\hat{\sigma}_{2,R}^{FT} \equiv \hat{\sigma}_{2,G^*}^{FT}$, $\hat{\sigma}_{2,R}^{P} \equiv \hat{\sigma}_{2,G^*}^{P}$, $\hat{\sigma}_{3,R}^{FT} \equiv \hat{\sigma}_{3,G^*}^{FT}$ and $\hat{\sigma}_{3,R}^{P} \equiv \hat{\sigma}_{3}^{P}$ where the latter sets are defined in our baseline model. With this we then have

$$\begin{aligned} L(V_R) &= \sum_{s \in \hat{\sigma}_{2,R}^{FT} \cup \hat{\sigma}_{2,R}^{P}} p(s)\left[qk(s)\left|\Gamma(s)\right| + c(s) + c^{*}(s)\right] \\ &+ \sum_{s \in \hat{\sigma}_{3,R}^{FT} \cup \hat{\sigma}_{3,R}^{P}} p(s)\left|\Gamma(s)\right|. \end{aligned}$$

4.1.1. Cash versus Retaliation

By comparing $L(V_R)$ and $L(V_C)$, we can characterize the conditions under which the V_R institution is preferred to the V_C institution. To build intuition, it is useful to start with the extreme case in which $m(s) = 0$ so that damages are always correctly assessed (but the court still makes the wrong ruling with positive probability). It is straightforward to confirm that all thresholds are then exactly the same under both institutions with the exception that $\mu_{2,C}^P(s) \equiv \frac{\gamma_G(s)}{|\gamma_{G^*}^*(s)|}\mu_{2,R}^P(s)$, which implies that $\mu_{2,C}^P(s) > \mu_{2,R}^P(s)$ since $\gamma_G(s) - |\gamma_{G^*}^*(s)| > 0$ in $s \in \sigma^P$, where to aid comparison we use the notation $\mu_{2,R}^P(s) \equiv \mu_2^P(s)$ with the latter threshold defined in our baseline model. Defining the set $\Omega_{2,C}^P \equiv \left\{s \in \sigma^{FT} \mid qk(s) \in \left[\mu_{2,R}^P(s), \mu_{2,C}^P(s)\right]\right\}$, we can thus write:

$$\begin{aligned} L(V_R) - L(V_C) &= \sum_{s \in \hat{\sigma}_{2,C}^P} p(s)\, qk(s)\, |\Gamma(s)| \\ &\quad + \sum_{s \in \Omega_{2,C}^P} p(s)\left\{|\Gamma(s)| - [c(s) + c^*(s)]\right\} \\ &> 0, \end{aligned}$$

where the term on the second line corresponds to the efficiency gain from V_C associated with the set $\Omega_{2,C}^P$ discussed above and the term on the first line reflects the efficiency gain from V_C in the set of states $\hat{\sigma}_{2,R}^P$ that stem from the accurate court-assessed damages that allow Home to pay these damages and continue to choose $\tau = P$ when the court mistakenly rules FT. Hence, if cash damages are assessed with perfect accuracy, the V_C institution unambiguously dominates the V_R institution ($L(V_R) - L(V_C) > 0$). In essence, under the V_C institution with perfect court damage assessments Home can continue to choose $\tau = P$ in states $s \in \sigma^P$ regardless of the ruling, which implies more efficient policy choices in states $s \in \hat{\sigma}_{2,R}^P \cup \Omega_{2,C}^P$.[27]

The discussion above makes clear that the V_R institution can only be optimal if cash damages are assessed with sufficient inaccuracy. We therefore now consider the difference $L(V_R) - L(V_C)$ for the general case in which $m(s) > 0$. For this case, it is easy to verify that $\mu_{1,C}^{FT}(s) < \mu_{1,R}^{FT}(s)$, $\mu_{2,C}^{FT}(s) < \mu_{2,R}^{FT}(s)$, $\mu_{1,C}^P(s) > \mu_{1,R}^P(s)$, and $\mu_{2,C}^P(s) > \mu_{2,R}^P(s)$, where again to aid comparison we use the notation $\mu_{1,R}^{FT}(s) \equiv \mu_1^{FT}(s)$, $\mu_{2,R}^{FT}(s) \equiv \mu_{2,G^*}^{FT}(s)$ and $\mu_{1,R}^P(s) \equiv \mu_{1,G^*}^P(s)$, with the latter thresholds defined in our baseline model. Defining the sets $\Omega_{1,C}^{FT}$, $\Omega_{2,C}^{FT}$, and $\Omega_{1,C}^P$, analogously

[27] Notice, however, that even when cash damages are assessed with perfect accuracy, the V_C institution still delivers the inefficient policy choice $\tau = FT$ in states $s \in \hat{\sigma}_{3,C}^P$ where court quality is so bad that the court is not even invoked.

to the set $\Omega_{2,C}^{P}$ above, we can then write:

$$\begin{aligned} L(V_R) - L(V_C) &= -\sum_{s\in\hat{\sigma}_{2,C}^{FT}} p(s)\left[1-qk(s)\right]m(s)\left|\Gamma(s)\right| \\ &+ \sum_{s\in\hat{\sigma}_{2,C}^{P}} p(s)\,qk(s)\left[1-m(s)\right]\left|\Gamma(s)\right| \\ &- \sum_{s\in\Omega_{1,C}^{FT}} p(s)\left[qk(s)\left|\Gamma(s)\right| + c(s) + c^{*}(s)\right] \\ &- \sum_{s\in\Omega_{2,C}^{FT}} p(s)\left\{\left[1-qk(s)\right]\left|\Gamma(s)\right| + c(s) + c^{*}(s)\right\} \\ &+ \sum_{s\in\Omega_{1,C}^{P}} p(s)\left\{qk(s)\left|\Gamma(s)\right| + c(s) + c^{*}(s)\right\} \\ &+ \sum_{s\in\Omega_{2,C}^{P}} p(s)\left\{1-\left[qk(s)\right]\left|\Gamma(s)\right| + c(s) + c^{*}(s)\right\} \end{aligned} \tag{4.3}$$

According to (4.3), the V_R institution can only be optimal ($L(V_R) - L(V_C) < 0$) under two conditions. First, the court has to be sufficiently bad at assessing cash damages relative to making rulings ($m(s)$ has to be high relative to $qk(s)$). As we have noted above, if $m(s)$ is sufficiently close to zero the V_C institution must dominates the V_R institution; and court ruling errors favor the V_C institution because when those errors occur Home can pay damages and avoid inefficient policy choices. And second, it has to be sufficiently likely that free trade is the efficient trade policy. This second condition, which is most obvious in the extreme case in which the set $s \in \sigma^P$ is empty so that only the negative terms in the above expression remain, reflects the fact that mistakes in the assessment of cash damages can effectively overturn the court's ruling of FT, which is bad if FT is the right ruling (i.e., in σ^{FT}) but good if FT is the wrong ruling (i.e., in σ^P). Under these conditions, the first term in (4.3) is large relative to the second term in absolute value and can turn the entire expression negative. Intuitively, if the court is very good at making rulings, cash damages only become an issue in states $s \in \sigma^{FT}$, since it will typically confirm Home's choice of $\tau = P$ in states $s \in \sigma^P$. But in states $s \in \sigma^{FT}$, cash damages can only overturn an efficient ruling and are more likely to do so the higher the probability that they are too low. The terms on the four remaining lines of (4.3) then capture the fact that mistakes in assessing damages make it more likely that Home continues to choose $\tau = P$, which is bad in states $s \in \sigma^{FT}$ and good in states $s \in \sigma^P$.

We summarize these results in:

Proposition 4. *Allowing for retaliation instead of cash damages in a trade agreement is optimal if (1) the court's ability to assess cash damages is sufficiently bad relative to its ability to assess the state of the world, and (2) free trade is sufficiently likely to be the efficient policy choice.*

4.2. Investment Agreements

We now consider the choice of retaliation versus cash payments in the context of BITs. As before, we build on the findings of the previous section and assume that the foreign investor has standing in investment disputes. And as before, to make our points as clearly as possible we assume that (i) when retaliation is the remedy, retaliation is so inefficient and costly that the host government always switches to $\iota = FT$ when convicted by the DSB to avoid retaliation, and (ii) when cash payments are the remedy, cash payments are perfectly efficient so that consumer surplus and producer surplus can be costlessly transferred internationally. As before, (i) implies that the analysis of BITs with retaliation is exactly the same as the earlier analysis of BITs in our baseline model.

To derive conditions under which cash damages are optimal in a BIT, we again need to first derive the equilibrium actions when cash damages are available, keeping in mind that the equilibrium actions under retaliation remain unchanged from our baseline model. With the possibility of cash damages, Foreign now files a complaint if $\iota = T$ and

$$\Pr\left(\text{ruling is } FT \mid s\right)\left\{\Pr\left(\iota = FT\right)\left|\gamma^*_{I^*}\left(I^*\right)\right| + \Pr\left(\iota = T\right) E\left[d^*_{I^*}\left(I^*\right) \mid \iota = T\right]\right\} > c^*\left(I^*, s\right). \tag{4.4}$$

If the host government is constrained, it chooses $\iota = \iota_{FB}$. Otherwise, it chooses $\iota = T$ if either the above condition is violated or

$$\begin{aligned} &\Pr\left(\text{ruling is } T \mid s\right)\gamma_G\left(I^*, s\right) + \\ &\Pr\left(\text{ruling is } FT \mid s\right)\Pr\left(\iota = T\right)\left\{\gamma_G\left(I^*, s\right) - E\left[d^*_{I^*}\left(I^*\right) \mid \iota = T\right]\right\} > c\left(I^*, s\right) \end{aligned} \tag{4.5}$$

Analogous to above, here $\Pr\left(\iota = FT\right)$ is the probability that the host government chooses to cease and desist in the face of a court ruling of FT while $\Pr\left(\iota = T\right)$ is the probability that Home instead continues with $\iota = T$ and pays damages, so that $\Pr\left(\iota = FT\right) = 1 - \Pr\left(\iota = T\right)$. Notice that the court damage assessment $d^*_{I^*}\left(I^*\right)$ now refers to the damages suffered by the foreign investor rather than the foreign government as was the case in our analysis of trade agreements.

We can now derive the equilibrium actions for each state s. Defining the thresholds $\bar{\mu}_{1,C}^{FT}(s) \equiv \frac{c\epsilon(s)\gamma_G(I^*,s)-m(s)\left[\gamma_G(I^*,s)-\underline{d}_{I^*}^*\right]}{\gamma_G(I^*,s)-m(s)\left[\gamma_G(I^*,s)-\underline{d}_{I^*}^*\right]}$, $\bar{\mu}_{2,C}^{FT}(s) \equiv 1-\frac{c^*\epsilon^*(s)PS(I^*)}{[1-2m(s)]PS(I^*)+m(s)\underline{d}_{I^*}^*}$, $\bar{\mu}_{1,C}^{T}(s) \equiv \frac{c^*\epsilon^*(s)PS(I^*)}{m(s)PS(I^*)+[1-2m(s)]PS(I^*)+m(s)\underline{d}_{I^*}^*}$, $\bar{\mu}_{2,C}^{T}(s) \equiv \frac{\gamma_G(I^*,s)-c\epsilon(s)\gamma_G(I^*,s)}{\gamma_G(I^*,s)-\left\{[1-2m(s)][\gamma_G(I^*,s)-PS(I^*)]+m(s)\left[\gamma_G(I^*,s)-\underline{d}_{I^*}^*\right]\right\}}$, and noting as before that $\bar{\mu}_{1,C}^{FT}(s) < \bar{\mu}_{2,C}^{FT}(s)$ and $\bar{\mu}_{1,C}^{P}(s) < \bar{\mu}_{2,C}^{P}(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (4.4) and (4.5) immediately imply the following result:

Lemma 5. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\bar{\mu}_{1,C}^{FT}(s) < \bar{\mu}_{2,C}^{FT}(s)$ and $\bar{\mu}_{1,C}^{P}(s) < \bar{\mu}_{2,C}^{P}(s)$, equilibrium actions under cash damages can be characterized as follows:*

1. *In states $s \in \sigma^{FT}$:*

    1. *If the host government is constrained: We have $\iota = FT$ and no dispute.*
    2. *If the host government is unconstrained: If DSB quality is high in the sense that $qk(s) < \bar{\mu}_{1,C}^{FT}(s)$, we have $\iota = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\bar{\mu}_{1,C}^{FT}(s), \bar{\mu}_{2,C}^{FT}(s)\right]$, we have $\iota = T$ and a dispute; if DSB quality is low in the sense that $qk(s) > \bar{\mu}_{2,C}^{FT}(s)$, we have $\iota = T$ and no dispute.*

2. *In states $s \in \sigma^{T}$:*

    1. *If the host government is constrained: We have $\iota = T$, no dispute if $qk(s) < \bar{\mu}_{1,C}^{T}(s)$, and a dispute if $qk(s) > \bar{\mu}_{1,C}^{T}(s)$.*
    2. *If the host government is unconstrained: If DSB quality is high in the sense that $qk(s) < \bar{\mu}_{1,C}^{T}(s)$, we have $\iota = T$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\bar{\mu}_{1,C}^{T}(s), \bar{\mu}_{2,C}^{T}(s)\right]$, we have $\iota = T$ and a dispute; if DSB quality is low in the sense that $qk(s) > \bar{\mu}_{2,C}^{T}(s)$, we have $\iota = FT$ and no dispute.*

For future reference, we denote the different action sets by $\bar{\sigma}_{1,C}^{FT} \equiv \left\{s \in \bar{\sigma}^{FT} \mid qk(s) < \bar{\mu}_{1,C}^{FT}(s)\right\}$, $\bar{\sigma}_{2,C}^{FT} \equiv \left\{s \in \bar{\sigma}^{FT} \mid qk(s) \in \left[\bar{\mu}_{1,C}^{FT}(s), \bar{\mu}_{2,C}^{FT}(s)\right]\right\}$, and $\bar{\sigma}_{3,C}^{FT} \equiv \left\{s \in \bar{\sigma}^{FT} \mid qk(s) > \bar{\mu}_{2,C}^{FT}(s)\right\}$, as well as $\bar{\sigma}_{1,C}^{T} \equiv \left\{s \in \bar{\sigma}^{T} \mid qk(s) < \bar{\mu}_{1,C}^{T}(s)\right\}$, $\bar{\sigma}_{2,C}^{T} \equiv \left\{s \in \bar{\sigma}^{T} \mid qk(s) \in \left[\bar{\mu}_{1,C}^{T}(s), \bar{\mu}_{2,C}^{T}(s)\right]\right\}$, and, $\bar{\sigma}_{3,C}^{T} \equiv \left\{s \in \bar{\sigma}^{T} \mid qk(s) > \bar{\mu}_{2,C}^{T}(s)\right\}$.

We can now characterize the ex-ante returns to foreign investors under cash damages for a given level of foreign investment I^*, which we denote by $E_s\left[\rho_{I^*}^{*C}\left(I^*,s\right)\right]$:

$$\begin{aligned}
E_s\left[\rho_{I^*}^{*C}\left(I^*,s\right)\right] &= \left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{1,C}^{FT}\cup\bar{\sigma}_{3,C}^{T}} p\left(s\right)\tilde{P}\left(I^*\right) \\
&+\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{2,C}^{FT}} p\left(s\right)\left\{\left[1-qk\left(s\right)\right]\left\{m\left(s\right)\frac{\underline{d}_{I^*}^{*}}{I^*}+\left[1-m\left(s\right)\right]\tilde{P}\left(I^*\right)\right\}-\frac{c^*\left(I^*,s\right)}{I^*}\right\} \\
&+\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{2,C}^{T}} p\left(s\right)\left\{qk\left(s\right)\left[\left[1-m\left(s\right)\right]\tilde{P}\left(I^*\right)+m\left(s\right)\frac{\underline{d}_{I^*}^{*}}{I^*}\right]-\frac{c^*\left(I^*,s\right)}{I^*}\right\} \\
&+\bar{p}p^{FT}\tilde{P}\left(I^*\right) \\
&+\bar{p}\sum_{s\in\bar{\sigma}_{2,C}^{T}\cup\bar{\sigma}_{3,C}^{T}} p\left(s\right)qk\left(s\right)\left\{\left[1-2m\left(s\right)\right]\tilde{P}\left(I^*\right)+m\left(s\right)\frac{\underline{d}_{I^*}^{*}}{I^*}+m\left(s\right)\frac{\bar{d}_{I^*}^{*}}{I^*}-\frac{c^*\left(I^*,s\right)}{I^*}\right\}.
\end{aligned}$$

Keeping in mind the up-front investment incentives that the host government offers to the foreign investor to ensure a return at the world interest rate r^*, the ex-ante payoff to the host government under cash damages for a given level of foreign investment I^* is given by:

$$\begin{aligned}
&E_s\left[\omega_{I^*}^{C}\left(I^*,s\right)\right] \\
= \;&\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{1,C}^{FT}} p\left(s\right)CS\left(I^*\right) \\
&+\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{2,C}^{FT}} p\left(s\right)\left\{CS\left(I^*\right)+qk\left(s\right)\kappa PS\left(I^*\right)+\left[1-qk\left(s\right)\right]m\left(s\right)\left[\kappa PS\left(I^*\right)-\underline{d}_{I^*}^{*}\right]-c\left(I^*,s\right)\right\} \\
&+\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{3,C}^{FT}} p\left(s\right)\left[CS\left(I^*\right)+\kappa PS\left(I^*\right)\right] \\
&-\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{2,C}^{T}} p\left(s\right)\left\{qk\left(s\right)\left\{\left[1-2m\left(s\right)\right]PS\left(I^*\right)+m\left(s\right)\left[e\left(I^*,s\right)-CS\left(I^*\right)+\underline{d}_{I^*}^{*}\right]\right\}+c\left(I^*,s\right)\right\} \\
&+\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{3,C}^{T}} p\left(s\right)\left[CS\left(I^*\right)-e\left(I^*,s\right)\right]+\bar{p}p^{FT}CS\left(I^*\right) \\
&-\bar{p}\sum_{s\in\bar{\sigma}_{2,C}^{T}\cup\bar{\sigma}_{3,C}^{T}} p\left(s\right)\left\{qk\left(s\right)\left\{\left[1-2m\left(s\right)\right]PS\left(I^*\right)+m\left(s\right)\underline{d}_{I^*}^{*}+m\left(s\right)\bar{d}_{I^*}^{*}\right\}+c\left(I^*,s\right)\right\} \\
&-\left\{r^*-E_s\left[\rho_{I^*}^{*}\left(I^*,s\right)\right]\right\}I^*.
\end{aligned}$$

Substituting the earlier expression for $E_s\left[\rho_{I^*}^{*C}\left(I^*,s\right)\right]$ then yields:

$$\begin{aligned}
& E_s\left[\omega_{I^*}^{C}\left(I^*,s\right)\right] \\
= \; & \left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{1,C}^{FT}} p\left(s\right)\left[CS\left(I^*\right)+PS\left(I^*\right)\right] \\
& +\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{2,C}^{FT}} p\left(s\right)\left\{CS\left(I^*\right)+\left[1-\left(1-\kappa\right)qk\left(s\right)\right]PS\left(I^*\right)-c\left(I^*,s\right)-c^*\left(I^*,s\right)\right\} \\
& +\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{3,C}^{FT}} p\left(s\right)\left[CS\left(I^*\right)+\kappa PS\left(I^*\right)\right] \\
& -\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{2,C}^{T}} p\left(s\right)\left[c\left(I^*,s\right)+c^*\left(I^*,s\right)\right] \\
& +\left(1-\bar{p}\right)\sum_{s\in\bar{\sigma}_{3,C}^{T}} p\left(s\right)\left[CS\left(I^*\right)+PS\left(I^*\right)-e\left(I^*,s\right)\right] \\
& +\bar{p}p^{FT}\left[CS\left(I^*\right)+PS\left(I^*\right)\right] \\
& -\bar{p}\sum_{s\in\bar{\sigma}_{2,C}^{T}\cup\bar{\sigma}_{3,C}^{T}} p\left(s\right)\left[c\left(I^*,s\right)+c^*\left(I^*,s\right)\right] \\
& -r^*I^*.
\end{aligned}$$

4.2.1. Cash versus Retaliation

We now provide sufficient conditions for the host government's ex-ante expected welfare under a BIT with cash damages to exceed its ex-ante expected welfare under a BIT that relies on retaliation. We do so by comparing $E_s\left[\omega_{I^*}^{C}\left(I^*,s\right)\right]$ to $E_s\left[\omega_{I^*}^{R}\left(I^*,s\right)\right]$, where $E_s\left[\omega_{I^*}^{R}\left(I^*,s\right)\right]$ is the host government's ex-ante expected welfare under a BIT with retaliation and an ISDS and is given by (3.11) from the previous section, and where we evaluate these expressions at the optimal level of investment under a BIT with retaliation and an ISDS, which we characterized in (3.12) from the previous section and now denote by $\bar{I}^{*R}$. Clearly, if $E_s\left[\omega_{I^*}^{C}\left(\bar{I}^{*R},s\right)\right]\geq E_s\left[\omega_{I^*}^{R}\left(\bar{I}^{*R},s\right)\right]$, it must also be that $E_s\left[\omega_{I^*}^{C}\left(\bar{I}^{*C},s\right)\right]\geq E_s\left[\omega_{I^*}^{R}\left(\bar{I}^{*R},s\right)\right]$, where $\bar{I}^{*C}$ is the level of investment optimal under a BIT with cash damages and an ISDS.

We begin again by considering the special case in which the court is perfectly accurate at assessing cash damages ($m\left(s\right)=0$). It is straightforward to see that all thresholds are exactly the same under both institutions in this special case with the only exception that $\bar{\mu}_{2,C}^{T}\left(s\right)\equiv\frac{\gamma_G(I^*,s)}{PS(I^*)}\bar{\mu}_{2,R}^{T}\left(s\right)$, which implies that $\bar{\mu}_{2,C}^{T}\left(s\right)>\bar{\mu}_{2,R}^{T}\left(s\right)$ since $\gamma_G\left(I^*,s\right)-PS\left(I^*\right)>0$ in $s\in\sigma^T$, where to aid comparison we use the notation $\bar{\mu}_{2,R}^{T}\left(s\right)\equiv\bar{\mu}_{2}^{T}\left(s\right)$ with the latter threshold

defined in our baseline model. Defining the set $\bar{\Omega}_{2,C}^{T} \equiv \left\{ s \in \sigma^{FT} \mid qk\left(s\right) \in \left[\bar{\mu}_{2,R}^{T}\left(s\right), \bar{\mu}_{2,C}^{T}\left(s\right) \right] \right\}$ and defining $\Delta_C\left(I^*, s\right) \equiv E_s\left[\omega_{I^*}^{C}\left(I^*, s\right)\right] - E_s\left[\omega_{I^*}^{R}\left(I^*, s\right)\right]$, we can thus write:

$$\begin{aligned}
\Delta_C\left(\bar{I}^{*R}, s\right) &= (1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,C}^{T}} p\left(s\right) qk\left(s\right) \left[e\left(\bar{I}^{*R}, s\right) - CS\left(\bar{I}^{*R}\right) - PS\left(\bar{I}^{*R}\right) \right] \\
&\quad + (1-\bar{p}) \sum_{s \in \bar{\Omega}_{2,C}^{T}} p\left(s\right) \left[e\left(\bar{I}^{*R}, s\right) - CS\left(\bar{I}^{*R}\right) - PS\left(\bar{I}^{*R}\right) - c\left(\bar{I}^{*R}, s\right) - c^*\left(\bar{I}^{*R}, s\right) \right] \\
&\quad + \bar{p} \sum_{s \in \bar{\sigma}_{2,C}^{T} \cup \bar{\sigma}_{3,C}^{T}} p\left(s\right) qk\left(s\right) \left[e\left(\bar{I}^{*R}, s\right) - CS\left(\bar{I}^{*R}\right) - PS\left(\bar{I}^{*R}\right) \right] \\
&> 0
\end{aligned}$$

Hence, as with trade agreements, the use of cash damages in a BIT unambiguously dominates the use of retaliation if cash damages are assessed with perfect accuracy. Notice also that the use of cash damages has benefits whether or not the host government can commit domestically. In the no-commitment case, the host government can continue with $\iota = T$ in states $s \in \sigma^T$ regardless of the ruling, which implies more efficient policy choices in states $s \in \bar{\sigma}_{2,C}^{T} \cup \bar{\Omega}_{2,C}^{T}$. In the commitment case, the host government can also continue with $\iota = T$ in states $s \in \sigma^T$ regardless of the ruling, which implies more efficient policy choices in states $\bar{\sigma}_{2,C}^{T} \cup \bar{\sigma}_{3,C}^{T}$.

For the general case $m\left(s\right) \geq 0$, it is easy to verify using the same notation convention as above that $\bar{\mu}_{1,C}^{FT}\left(s\right) < \bar{\mu}_{1,R}^{FT}\left(s\right)$, $\bar{\mu}_{2,C}^{FT}\left(s\right) < \bar{\mu}_{2,R}^{FT}\left(s\right)$, $\bar{\mu}_{1,C}^{T}\left(s\right) > \bar{\mu}_{1,R}^{T}\left(s\right)$, and $\bar{\mu}_{2,C}^{T}\left(s\right) > \bar{\mu}_{2,R}^{T}\left(s\right)$. Defining the sets $\bar{\Omega}_{1,C}^{FT}$, $\bar{\Omega}_{2,C}^{FT}$, and $\bar{\Omega}_{1,C}^{T}$, analogously to the set $\bar{\Omega}_{2,C}^{T}$ above, we can then write:

$$\begin{aligned}
\Delta_C\left(\bar{I}^{*R}, s\right) &= -(1-\bar{p}) \sum_{s \in \bar{\Omega}_{1,C}^{FT}} p\left(s\right) \left[(1-\kappa) qk\left(s\right) PS\left(\bar{I}^{*R}\right) + c\left(I^*, s\right) + c^*\left(\bar{I}^{*R}, s\right) \right] \qquad (4.6) \\
&\quad - (1-\bar{p}) \sum_{s \in \bar{\Omega}_{2,C}^{FT}} p\left(s\right) \left\{ \left[1 - qk\left(s\right)\right] (1-\kappa) PS\left(\bar{I}^{*R}\right) - c\left(\bar{I}^{*R}, s\right) - c^*\left(\bar{I}^{*R}, s\right) \right\} \\
&\quad + (1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,C}^{T}} p\left(s\right) qk\left(s\right) \left[e\left(\bar{I}^{*R}, s\right) - CS\left(\bar{I}^{*R}\right) - PS\left(\bar{I}^{*R}\right) \right] \\
&\quad + (1-\bar{p}) \sum_{s \in \bar{\Omega}_{1,C}^{T}} p\left(s\right) \left[c_{I^*}^{*}\left(\bar{I}^{*R}, s\right) + c\left(\bar{I}^{*R}, s\right) \right] \\
&\quad + (1-\bar{p}) \sum_{s \in \bar{\Omega}_{2,C}^{T}} p\left(s\right) \left[e\left(\bar{I}^{*R}, s\right) - CS\left(\bar{I}^{*R}\right) - PS\left(\bar{I}^{*R}\right) - c\left(\bar{I}^{*R}, s\right) - c^*\left(\bar{I}^{*R}, s\right) \right] \\
&\quad + \bar{p} \sum_{s \in \bar{\sigma}_{2,C}^{T} \cup \bar{\sigma}_{3,C}^{T}} p\left(s\right) qk\left(s\right) \left[e\left(\bar{I}^{*R}, s\right) - CS\left(\bar{I}^{*R}\right) - PS\left(\bar{I}^{*R}\right) \right] \\
&\quad + \bar{p} \sum_{s \in \bar{\Omega}_{1,C}^{T}} p\left(s\right) \left[c\left(\bar{I}^{*R}, s\right) + c^*\left(\bar{I}^{*R}, s\right) \right].
\end{aligned}$$

According to (4.6), $\Delta_C\left(\bar{I}^{*R}, s\right) > 0$ and the inclusion of cash damages in a BIT with an ISDS is optimal, provided that two conditions are met. First, the court has to be sufficiently *good* at assessing cash damages relative to making rulings ($m(s)$ has to be low relative to $qk(s)$). If $m(s)$ is small, the sets $\bar{\Omega}_{1,C}^{FT}$ and $\bar{\Omega}_{2,C}^{FT}$ become smaller, and with them the only two negative terms in (4.6). And if $qk(s)$ is large, the effects of cash damages loom larger in states $s \in \sigma^T$ – where those effects are positive – since the host government is able to pay damages and continue with $\iota = T$ in states $s \in \sigma^T$ unless the assessed damages are prohibitively high, whereas the court often then also mistakenly authorizes $\iota = T$ in states $s \in \sigma^{FT}$ but damages are irrelevant in those states. And second, there has to be a non-trivial probability that a taking is the efficient policy. This is most obvious in the extreme case in which the set $s \in \sigma^T$ is empty so that only the negative terms in (4.6) remain. Just as in the case of trade agreements, the intuition is that mistakes in the assessment of cash damages can effectively overturn the court's ruling of FT, which is bad if FT is the right ruling (i.e., in σ^{FT}) but good if FT is the wrong ruling (i.e., in σ^T).

We summarize these results in:

Proposition 5. *Allowing for cash damages instead of retaliation in a BIT is optimal if (1) the court's ability to assess cash damages is sufficiently good relative to its ability to assess the state of the world, and (2) there is a non-trivial probability that a taking is the efficient policy.*

5. The Remedial Period

Thus far we have assumed that litigation is effectively instantaneous. When a case is filed, it is adjudicated immediately and the parties comply with the ruling immediately; hence, there is no possibility of any "pre-compliance harm" to exporters or investors. In this section, we augment our baseline models from section 3 to allow for the possibility of pre-compliance harm, and consider in the context of both trade agreements and investment agreements the following question: What is the optimal "remedial period"? That is, is it optimal to include retrospective damages in the remedy (damages for harm suffered before the case is finally adjudicated), or just prospective damages (damages that would arise after adjudication if the ruling is not obeyed)?

To capture these alternative remedy design possibilities, we let $\delta \in [0, 1]$ parameterize the fraction of the harm from the policy action at issue that occurs retrospectively, that is, prior to the court ruling. We interpret δ broadly: we have in mind all the costs of delay in adjudication

including lost profits on export sales, losses experienced due to impairment of sunk investments, and so on. If $\delta = 0$, there is no pre-ruling harm, as in section 3; at the other extreme, if $\delta = 1$, the harm has all occurred and is a bygone by the time of the ruling.

We think of lower values of δ as reasonable for the case of trade disputes, where the pre-ruling harm is attributable mostly to delay in securing market access, thus resulting in some lost profits if exports must be diverted to other markets temporarily, and so in our analysis of trade agreements we will highlight model results with that parameter range in mind. By contrast, we think of higher values of δ as reasonable for the case of investment disputes, where the pre-ruling harm is usually attributable to policy actions that diminish or destroy the returns to a sunk investment with a finite lifespan, and so in our analysis of BITs we will highlight model results with that parameter range in mind. For shorthand, throughout this section we will refer to δ as the degree of "litigation delay," though as discussed above we interpret δ broadly.

We build on our analysis in section 3, and assume that the trade agreement has adopted SSDS while the BIT has adopted ISDS. And to keep the comparison clean, we continue to assume that the prospective remedy for both trade agreements and BITs is a cease and desist order, just as in the models of section 3: hence, in the case where $\delta = 0$ and there is no pre-ruling harm, the augmented models that we develop in this section collapse to the original models of section 3. And finally, building on our analysis in section 4 we assume for the retrospective remedy that the trade agreement relies on retaliation while the BIT employs cash damages. In this section we take these features as exogenously given so that we may focus on the question of the optimal remedial period.[28]

5.1. Trade Agreements

Prospective damages Consider first the case of prospective damages. For $\delta > 0$, there are two changes implied for a trade agreement with SSDS relative to our earlier analysis of trade agreements in section 3.1 (i.e., relative to the case where $\delta = 0$). First, the conditions describing equilibrium behavior of the two governments are altered. And second, the joint surplus in any state conditional on government behavior is altered. We consider each in turn.

Consider first the foreign government's filing behavior under prospective damages. The

[28] As we note, our assumption that for the retrospective remedy the trade agreement relies on retaliation while the BIT relies on cash damages can be rationalized by our findings in section 4, but there are also other arguments that can provide support for this assumption (see, for example, Sykes, 2005, Limao and Saggi, 2008, and Bagwell and Staiger, 2010, note 10).

foreign government files a complaint if and only if $\tau = P$ and its expected benefit of filing exceeds its cost of filing, that is

$$\Pr(\text{DSB ruling is } FT \mid s) \times [1-\delta]|\gamma^*_{G^*}(s)| > c^*(s). \tag{5.1}$$

Condition (5.1) is the "filing" condition for the foreign government to invoke the DSB in response to a policy choice by the importer government of $\tau = P$.

Next consider the importer government's policy choice under prospective damages. This government chooses $\tau = P$ if either (5.1) fails – because then the importer government can set $\tau = P$ without triggering a dispute – or if (5.1) holds and the expected benefit to the importer government from trade protection exceeds the cost to the importer government of a dispute:

$$\delta\gamma_G(s) + \Pr(\text{DSB ruling is } P \mid s) \times [1-\delta]\gamma_G(s) > c(s)$$

or

$$\Pr(\text{DSB ruling is } FT \mid s) \times \delta\gamma_G(s) + \Pr(\text{DSB ruling is } P \mid s) \times \gamma_G(s) > c(s). \tag{5.2}$$

We can now derive the equilibrium actions for each state s. Again in what follows we assume that the states where the vague contract is unambiguous are measure zero, so we can focus only on states where the court if invoked must interpret the contract. Defining the thresholds $\mu^{FT}_{1,Pro}(s) \equiv \frac{c(s)-\delta\gamma_G(s)}{[1-\delta]\gamma_G(s)}$, $\mu^{FT}_{2,Pro}(s) \equiv 1 - \frac{c^*(s)}{[1-\delta]|\gamma^*_{G^*}(s)|}$, $\mu^{P}_{1,Pro}(s) \equiv \frac{c^*(s)}{[1-\delta]|\gamma^*_{G^*}(s)|}$, and $\mu^{P}_{2,Pro}(s) \equiv (1 - \frac{c(s)-\delta\gamma_G(s)}{[1-\delta]\gamma_G(s)})$, and noting that $\mu^{FT}_{1,Pro}(s) < \mu^{FT}_{2,Pro}(s)$ and $\mu^{P}_{1,Pro}(s) < \mu^{P}_{2,Pro}(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (5.1) and (5.2) immediately imply the following result:

Lemma 6. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\mu^{FT}_{1,Pro}(s) < \mu^{FT}_{2,Pro}(s)$ and $\mu^{P}_{1,Pro}(s) < \mu^{P}_{2,Pro}(s)$, equilibrium actions can be characterized as follows:*

1. *In states $s \in \sigma^{FT}$: If DSB quality is high in the sense that $qk(s) < \mu^{FT}_{1,Pro}(s)$, we have $\tau = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in [\mu^{FT}_{1,Pro}(s), \mu^{FT}_{2,Pro}(s)]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu^{FT}_{2,Pro}(s)$, we have $\tau = P$ and no dispute.*

2. *In states $s \in \sigma^P$: If DSB quality is high in the sense that $qk(s) < \mu^P_{1,Pro}(s)$, we have $\tau = P$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\mu^P_{1,Pro}(s), \mu^P_{2,Pro}(s)\right]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu^P_{2,Pro}(s)$, we have $\tau = FT$ and no dispute.*

Note that if $\delta = 0$ and there is hence no litigation delay, the above characterization of equilibrium behavior collapses to our earlier section 3.1 analysis of trade agreements under SSDS. On the other hand, if δ is sufficiently close to 1, (5.1) and (5.2) together with our focus on the relatively-low-dispute-cost case imply that the importer government will always choose P and the exporter government will never invoke the DSB, and hence for δ in this range the trade agreement with prospective damages delivers the noncooperative outcome and therefore becomes valueless. Formally, the critical level of δ beyond which a trade agreement with prospective damages would be valueless, which we denote by $\hat{\delta}$, is defined by

$$\hat{\delta} = 1 - \min_s \left[\frac{c^*(s)}{|\gamma^*_{G^*}(s)|}\right] > 0$$

where the inequality follows from our focus on the relatively-low-dispute-cost case. For $\delta \in [\hat{\delta}, 1]$, it follows from the above characterization of equilibrium behavior that in all states $s \in \sigma^{FT}$ the importer government chooses P and the exporter government will not file (because $qk(s) > 0 \geq \mu^{FT}_{2,Pro}(s)$ for all s), and likewise in all states $s \in \sigma^P$ the importer government chooses P and the exporter government does not file (because $qk(s) < 1 \leq \mu^P_{1,Pro}(s)$ for all s).

Next we write down the expected efficiency loss, relative to the first-best outcome, for a trade agreement with SSDS under prospective damages when the litigation delay is given by δ. For the low-relative-dispute-cost case on which we focus, this loss, which we denote by $L(V_{Pro},)$, is given by:

$$\begin{aligned} L(V_{Pro}, \delta) \;=\; & \sum_{s \in \hat{\sigma}^{FT}_{Pro,2}(\delta)} p(s)\delta|\Gamma(s)| + \\ & \sum_{s \in \{\hat{\sigma}^{FT}_{Pro,2}(\delta) \cup \hat{\sigma}^{P}_{Pro,2}(\delta)\}} p(s)\{[1-\delta)]qk(s)|\Gamma(s)| + [c(s) + c^*_f(s)]\} + \\ & \sum_{s \in \{\hat{\sigma}^{FT}_{Pro,3}(\delta) \cup \hat{\sigma}^{P}_{Pro,3}(\delta)\}} p(s)|\Gamma(s)|. \end{aligned} \tag{5.3}$$

Here, $\hat{\sigma}^{FT}_{Pro,2}(\delta)$ denotes the set of states for which FT is efficient, the importer government chooses P, and a complaint is filed (i.e., s such that $s \in \sigma^{FT}$ and $qk(s) \in [\mu^{FT}_{1,Pro}(s), \mu^{FT}_{2,Pro}(s)]$).

Similarly, $\hat{\sigma}^{P}_{Pro,2}(\delta)$ denotes the set of states for which P is efficient, the importer government chooses P, and a complaint is filed (i.e., s such that $s \in \sigma^{P}$ and $qk(s) \in [\mu^{P}_{1,Pro}(s), \mu^{P}_{2,Pro}(s)]$). And finally, $\hat{\sigma}^{FT}_{Pro,3}(\delta)$ denotes the set of states for which FT is efficient, the importer government chooses P, and no complaint is filed (i.e., s such that $s \in \sigma^{FT}$ and $qk(s) > \mu^{FT}_{2,Pro}(s)$), while $\hat{\sigma}^{P}_{Pro,3}(\delta)$ denotes the set of states for which P is efficient, the importer government chooses FT, and no complaint is filed (i.e., s such that $s \in \sigma^{P}$ and $qk(s) > \mu^{P}_{2,Pro}(s)$).

The broad properties of $L(V_{Pro}, \delta)$ as a function of δ are clear. For $\delta = 0$ we have

$$L(V_{Pro}, \delta = 0) = \sum_{s \in \{\hat{\sigma}^{FT}_{G^*,2} \cup \hat{\sigma}^{P}_{G^*,2}\}} p(s)\{qk(s)|\Gamma(s)| + [c(s) + c^{*}_{f}(s)]\} + \sum_{s \in \{\hat{\sigma}^{FT}_{G^*,3} \cup \hat{\sigma}^{P}_{3}\}} p(s)|\Gamma(s)|.$$

Therefore, by (3.3), $L(V_{Pro}, \delta = 0)$ is equal to $L(V_{G^*})$, the loss relative to the first best under a trade agreement with SSDS when there is no litigation delay as analyzed in section 3.1. And as we showed in section 3.1, provided that court quality is above a minimal threshold level this loss is less than the loss relative to the first best that would be experienced absent a trade agreement, when P would be chosen with impunity in all states; and if court quality is sufficiently high, this loss goes to zero and the first best is achieved. On the other hand, for $\delta \in [\hat{\delta}, 1]$ we have

$$L(V_{G^*_{Pro}}, \delta \geq \hat{\delta}) = \sum_{s \in \sigma^{FT}} p(s)|\Gamma(s)|,$$

which is equivalent to the loss relative to the first best that would be experienced absent a trade agreement.

Retrospective damages We consider next the case of retrospective damages. Under retrospective damages, if the DSB rules for FT, the importer government must both cease and desist its P policy and revert to FT henceforth (prospective damages), *and* it must make damage payments to the exporter government in the amount of the harm $\delta|\gamma^{*}_{G^*}(s)|$ already suffered (retrospective damages).[29]

A key question is the form that such retrospective damage payments take. As indicated above and consistent with GATT/WTO practice and our results from section 4, we assume that in the context of a trade dispute these damage payments take the form of additional tariff adjustments in other sectors, made either by the importer government or the exporter

[29] We do not include the litigation costs $c^{*}(s)$ borne by the exporter government in this damage payment, but this could be considered as well.

government, that amount to a *costly transfer* to the exporter government.[30] We capture the cost of such ex-post transfers in this setting with the parameter $\beta \in (0,1]$ representing the fraction of each dollar given up by the importer government that reaches the exporter government. Hence, a damage payment of $\delta|\gamma^*_{G^*}(s)|$ received by the exporter government costs the importer government $\frac{1}{\beta}[\delta|\gamma^*_{G^*}(s)|]$ in lost surplus. In the context of trade agreements we will highlight outcomes that arise in the absence of cash transfers where β is small.[31]

Adopting retrospective damages will have implications for the conditions describing equilibrium behavior of the two governments, and for the joint surplus in any state conditional on government behavior is altered. Again we consider each in turn.

Consider first the foreign government's filing behavior under retrospective damages. The foreign government files a complaint if and only if $\tau = P$ and its expected benefit of filing exceeds its cost of filing, that is

$$\Pr(\text{DSB ruling is } FT \mid s) \times |\gamma^*_{G^*}(s)| > c^*(s). \tag{5.4}$$

Condition (5.4) is the "filing" condition for the foreign government to invoke the DSB in response to a policy choice by the importer government of $\tau = P$.

Next consider the importer government's policy choice under retrospective damages. This government chooses $\tau = P$ if either (5.4) fails – because then the importer government can set $\tau = P$ without triggering a dispute – or if (5.4) holds and the expected benefit to the importer government from trade protection exceeds the cost to the importer government of a dispute:

$$\delta\{[\gamma_G(s) - \Pr(\text{DSB ruling is } FT \mid s) \times \frac{|\gamma^*_{G^*}(s)|}{\beta}\} + [1-\delta]\Pr(\text{DSB ruling is } P \mid s) \times \gamma_G(s) > c(s)$$

or

$$\Pr(\text{DSB ruling is } P \mid s) \times \gamma_G(s) + \Pr(\text{DSB ruling is } FT \mid s) \times \{\delta[\gamma_G(s) - \frac{|\gamma^*_{G^*}(s)|}{\beta}]\} > c(s). \tag{5.5}$$

We can now derive the equilibrium actions for each state s in the presence of retrospective damages. As above in what follows we assume that the states where the vague contract is

[30] Our implicit assumption is that trade agreements have resulted in the importer government's tariffs in other sectors being set at efficient levels, so that if the importer government were to make an adjustment (downward) in these tariffs to pay damages to the exporter government, these adjustments would have negative efficiency consequences, just as would be the case if the exporter government were to collect damage payments by raising its tariffs against the importer government.

[31] In contrast to section 4, here we are assuming that the court can perfectly assess the level of damages $|\gamma^*_{G^*}(s)|$, so that we can focus on the inefficiency of retaliation as a transfer mechanism through the parameter β.

unambiguous are measure zero, so we can focus only on states where the court if invoked must interpret the contract. Defining the thresholds $\mu^{FT}_{1,Retro}(s) \equiv \frac{c(s)-\delta\gamma_G(s)+\delta\frac{|\gamma^*_{G^*}(s)|}{\beta}}{[1-\delta]\gamma_G(s)+\delta\frac{|\gamma^*_{G^*}(s)|}{\beta}}$, $\mu^{FT}_{2,Retro}(s) \equiv 1-\frac{c^*(s)}{|\gamma^*_{G^*}(s)|}$, $\mu^{P}_{1,Retro}(s) \equiv \frac{c^*(s)}{|\gamma^*_{G^*}(s)|}$, and $\mu^{P}_{2,Retro}(s) \equiv (1-\frac{c(s)-\delta\gamma_G(s)+\delta\frac{|\gamma^*_{G^*}(s)|}{\beta}}{[1-\delta]\gamma_G(s)+\delta\frac{|\gamma^*_{G^*}(s)|}{\beta}})$, and noting that $\mu^{FT}_{1,Retro}(s) < \mu^{FT}_{2,Retro}(s)$ and $\mu^{P}_{1,Retro}(s) < \mu^{P}_{2,Retro}(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (5.4) and (5.5) immediately imply the following result:

Lemma 7. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\mu^{FT}_{1,Retro}(s) < \mu^{FT}_{2,Retro}(s)$ and $\mu^{P}_{1,Retro}(s) < \mu^{P}_{2,Retro}(s)$, equilibrium actions can be characterized as follows:*

1. *In states $s \in \sigma^{FT}$: If DSB quality is high in the sense that $qk(s) < \mu^{FT}_{1,Retro}(s)$, we have $\tau = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in [\mu^{FT}_{1,Retro}(s), \mu^{FT}_{2,Retro}(s)]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu^{FT}_{2,Retro}(s)$, we have $\tau = P$ and no dispute.*

2. *In states $s \in \sigma^{P}$: If DSB quality is high in the sense that $qk(s) < \mu^{P}_{1,Retro}(s)$, we have $\tau = P$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\mu^{P}_{1,Retro}(s), \mu^{P}_{2,Retro}(s)\right]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu^{P}_{2,Retro}(s)$, we have $\tau = FT$ and no dispute.*

Note that as β approaches 0, $\mu^{FT}_{1,Retro}(s)$ approaches 1 and $\mu^{P}_{2,Retro}(s)$ approaches 0, and the importer government will always choose FT to avoid any possibility of having to make costly transfer payments to the foreign exporter government for retroactive damages (because then $qk(s) < 1 = \mu^{FT}_{1,Retro}(s)$ for all $s \in \sigma^{FT}$ and $qk(s) > 0 = \mu^{P}_{2,Retro}(s)$ for all $s \in \sigma^{P}$).[32]

Next we write down the expected efficiency loss, relative to the first-best outcome, for a trade agreement with SSDS under retrospective damages. For the low-relative-dispute-cost

[32] Notice too that as we allow β to approach 0 we must also have $c^*(s)$ approaching 0 in order to maintain our "relatively small litigation cost" focus and ensure that $\mu^{FT}_{1,Retro}(s) < \mu^{FT}_{2,Retro}(s)$.

case on which we focus, this loss, which we denote by $L(V_{G^*_{Retro}})$, is given by:

$$
\begin{aligned}
L(V_{G^*_{Retro}}) \;=\; & \sum_{s\in\hat{\sigma}^{FT}_{G^*_{Retro},2}} p(s)\delta|\Gamma(s)| + \\
& \sum_{s\in\hat{\sigma}^{FT}_{G^*_{Retro},2}\cup\hat{\sigma}^{P}_{G^*_{Retro},2}} p(s)\{[1-\delta]qk(s)|\Gamma(s)| + [c(s)+c^*(s)]\} + \\
& \sum_{s\in\hat{\sigma}^{FT}_{G^*_{Retro},3}\cup\hat{\sigma}^{P}_{G^*_{Retro},3}} p(s)|\Gamma(s)| + \\
& \sum_{s\in\hat{\sigma}^{FT}_{G^*_{Retro},2}} p(s)\{\delta[1-qk(s)][\frac{1-\beta}{\beta}]|\gamma^*_{G^*}(s)|\} + \\
& \sum_{s\in\hat{\sigma}^{P}_{G^*_{Retro},2}} p(s)\{\delta qk(s)[\frac{1-\beta}{\beta}]|\gamma^*_{G^*}(s)|\}.
\end{aligned}
\tag{5.6}
$$

Here, $\hat{\sigma}^{FT}_{G^*_{Retro},2}$ denotes the set of states for which FT is efficient, the importer government chooses P, and a complaint is filed (i.e., s such that $s \in \sigma^{FT}$ and $qk(s) \in [\mu^{FT}_{1,Retro}(s), \mu^{FT}_{2,Retro}(s)]$). Similarly, $\hat{\sigma}^{P}_{G^*_{Retro},2}$ denotes the set of states for which P is efficient, the importer government chooses P, and a complaint is filed (i.e., s such that $s \in \sigma^{P}$ and $qk(s) \in [\mu^{P}_{1,Retro}(s), \mu^{P}_{2,Retro}(s)]$). And finally, $\hat{\sigma}^{FT}_{G^*_{Retro},3}$ denotes the set of states for which FT is efficient, the importer government chooses P, and no complaint is filed (i.e., s such that $s \in \sigma^{FT}$ and $qk(s) > \mu^{FT}_{2,Retro}(s)$), while $\hat{\sigma}^{P}_{G^*_{Retro},3}$ denotes the set of states for which P is efficient, the importer government chooses FT, and no complaint is filed (i.e., s such that $s \in \sigma^{P}$ and $qk(s) > \mu^{P}_{2,Retro}(s)$).

Using (5.6) and (5.3), we can now define $\Delta_{Retro,Pro} \equiv L(V_{G^*_{Retro}}) - L(V_{G^*_{Pro}})$ and look for parameter ranges under which $\Delta_{Retro,Pro}$ is positive, and in this way identify conditions under which prospective damages will be optimal in a trade agreement. To facilitate this evaluation, we define the following sets:

$$
\begin{aligned}
\Omega^{FT}_{2,low} &\equiv \{s \in \sigma^{FT} | qk(s) \in (\mu^{FT}_{1,Pro}(s), \mu^{FT}_{1,Retro}(s))\} \\
\Omega^{FT}_{2,high} &\equiv \{s \in \sigma^{FT} | qk(s) \in (\mu^{FT}_{2,Pro}(s), \mu^{FT}_{2,Retro}(s))\} \\
\Omega^{P}_{2,low} &\equiv \{s \in \sigma^{P} | qk(s) \in (\mu^{P}_{1,Retro}(s), \mu^{P}_{1,Pro}(s))\} \\
\Omega^{P}_{2,high} &\equiv \{s \in \sigma^{P} | qk(s) \in (\mu^{P}_{2,Retro}(s), \mu^{P}_{2,Pro}(s))\} \\
\Omega^{FT}_{3,low} &\equiv \{s \in \sigma^{FT} | qk(s) \in (\mu^{FT}_{2,Pro}(s), \mu^{FT}_{2,Retro}(s))\} \\
\Omega^{P}_{3,low} &\equiv \{s \in \sigma^{P} | qk(s) \in (\mu^{P}_{2,Retro}(s), \mu^{P}_{2,Pro}(s))\}.
\end{aligned}
\tag{5.7}
$$

We can now write

$$\begin{aligned}\Delta_{Retro,Pro} \equiv\ & L(V_{G^*_{Retro}}) - L(V_{G^*_{Pro}}) = \\ & \sum_{s\in\Omega^{FT}_{2,high}} p(s)\delta|\Gamma(s)| - \sum_{s\in\Omega^{FT}_{2,low}} p(s)\delta|\Gamma(s)| + \\ & \sum_{s\in\Omega^{FT}_{2,high}\cup\Omega^{P}_{2,low}} p(s)\{[1-\delta]qk(s)|\Gamma(s)| + [c(s)+c^*(s)]\} - \\ & \sum_{s\in\Omega^{FT}_{2,low}\cup\Omega^{P}_{2,high}} p(s)\{[1-\delta]qk(s)|\Gamma(s)| + [c(s)+c^*(s)]\} + \\ & \sum_{s\in\Omega^{P}_{3,low}} p(s)|\Gamma(s)| - \sum_{s\in\Omega^{FT}_{3,high}} p(s)|\Gamma(s)| + \\ & \sum_{s\in\hat{\sigma}^{FT}_{G^*_{Retro},2}} p(s)\{\delta[1-qk(s)][\frac{1-\beta}{\beta}]|\gamma^*_{G^*}(s)|\} + \\ & \sum_{s\in\hat{\sigma}^{P}_{G^*_{Retro},2}} p(s)\{\delta qk(s)[\frac{1-\beta}{\beta}]|\gamma^*_{G^*}(s)|\}.\end{aligned} \tag{5.8}$$

Using (5.8), it is straightforward to establish the following:

Proposition 6. *A prospective remedy is optimal for a trade agreement ($\Delta_{Retro,Pro}$ is positive) provided that the degree of litigation delay is sufficiently short (δ small), transfers in the context of a trade dispute are sufficiently costly (β small) and the quality of the court is sufficiently high (q low).*

Intuitively, for any fixed court quality q, the joint surplus under a trade agreement with retrospective damages will approach the joint surplus associated with FT in all states as β approaches zero and the cost of transfers becomes prohibitive. And if the quality of the court is fixed at a sufficiently high level, then as δ approaches zero so that litigation delay becomes sufficiently short, the joint surplus under a trade agreement with prospective damages can be brought arbitrarily close to the first best level, which exceeds the joint surplus associated with FT in all states and therefore beats a system with costly retrospective damages.

5.2. Investment Agreements

We now turn to the case of a BIT, and look for conditions under which retrospective damages would be optimal under the assumption that the BIT has adopted ISDS. Recall that, once

a level of FDI I^* has been sunk, our model of investment disputes in the context of a BIT works just like our model of trade disputes in the context of a trade agreement, with the host government now playing the role of the importer government from the previous subsection and the foreign investor now playing the role of the foreign exporter (but under ISDS, with standing) of the previous subsection.

Prospective damages Again we consider first the case of prospective damages. For $\delta > 0$, there are two changes implied for a BIT with ISDS relative to our earlier analysis of BITs in section 3.2 (i.e., relative to the case where $\delta = 0$). First, the conditions describing equilibrium behavior of the host government and the foreign investor are altered. And second, the payoffs to the host government and to the foreign investor in any state conditional on this equilibrium behavior are altered. We consider each in turn.

Consider first the foreign investor's filing behavior under prospective damages. The investor files a complaint if and only if $\iota = T$ and the expected benefit to the investor of filing exceeds its cost of filing, that is

$$\Pr(\text{DSB ruling is } FT \mid s) \times [1 - \delta]|\gamma^*_{I^*}(I^*)| > c^*(I^*, s). \tag{5.9}$$

Condition (5.9) is the foreign investor "filing" condition to invoke the DSB in response to a policy choice by the host government of $\iota = T$.

Next consider the host government's policy choice. When it has discretion to do so, this government chooses $\iota = T$ if either (5.9) fails – because then the host government can set $\iota = T$ without triggering a dispute – or if (5.9) holds and the expected benefit to the host government from a taking exceeds the cost to the host government of a dispute:

$$\delta\gamma_G(I^*, s) + \Pr(\text{DSB ruling is } T \mid s) \times [1 - \delta]\gamma_G(I^*, s) > c(I^*, s)$$

or

$$\Pr(\text{DSB ruling is } FT \mid s) \times \delta\gamma_G(I^*, s) + \Pr(\text{DSB ruling is } T \mid s) \times \gamma_G(I^*, s) > c(I^*, s). \tag{5.10}$$

We can now derive the equilibrium actions, conditional on investment level I^*, for each state s. As before, in what follows we assume that the states where the vague contract is unambiguous are measure zero, so we can focus only on states where the court if invoked must interpret the contract. Recall also from section 3.2 that the host government is constrained to implement

the first-best policies with probability $\bar{p}$ and can act at its own discretion with probability $1-\bar{p}$. Defining the thresholds $\bar{\mu}^{FT}_{1,Pro}(s) \equiv \frac{c\epsilon(s)-\delta}{[1-\delta]}$, $\bar{\mu}^{FT}_{2,Pro}(s) \equiv 1-\frac{c^*\epsilon^*(s)}{[1-\delta]}$, $\bar{\mu}^{T}_{1,Pro}(s) \equiv \frac{c^*\epsilon^*(s)}{[1-\delta]}$, and $\bar{\mu}^{T}_{2,Pro}(s) \equiv 1-\frac{c\epsilon(s)-\delta}{[1-\delta]}$, and noting as before that $\bar{\mu}^{FT}_{1,Pro}(s) < \bar{\mu}^{FT}_{2,Pro}(s)$ and $\bar{\mu}^{T}_{1,Pro}(s) < \bar{\mu}^{T}_{2,Pro}(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (3.9) and (3.10) immediately imply the following result:

Lemma 8. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\bar{\mu}^{FT}_{1,Pro}(s) < \bar{\mu}^{FT}_{2,Pro}(s)$ and $\bar{\mu}^{T}_{1,Pro}(s) < \bar{\mu}^{T}_{2,Pro}(s)$, equilibrium actions can be characterized as follows:*

1. *In states $s \in \sigma^{FT}$:*

    1. *If the host government is constrained: We have $\iota = FT$ and no dispute.*
    2. *If the host government is unconstrained: If DSB quality is high in the sense that $qk(s) < \bar{\mu}^{FT}_{1,Pro}(s)$, we have $\iota = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\bar{\mu}^{FT}_{1,Pro}(s), \bar{\mu}^{FT}_{2,Pro}(s)\right]$, we have $\iota = T$ and a dispute; if DSB quality is low in the sense that $qk(s) > \bar{\mu}^{FT}_{2,Pro}(s)$, we have $\iota = T$ and no dispute.*

2. *In states $s \in \sigma^{T}$:*

    1. *If the host government is constrained: We have $\iota = T$, no dispute if $qk(s) < \bar{\mu}^{T}_{1,Pro}(s)$, and a dispute if $qk(s) > \bar{\mu}^{T}_{1,Pro}(s)$.*
    2. *If the host government is unconstrained: If DSB quality is high in the sense that $qk(s) < \bar{\mu}^{T}_{1,Pro}(s)$, we have $\iota = T$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\bar{\mu}^{T}_{1,Pro}(s), \bar{\mu}^{T}_{2,Pro}(s)\right]$, we have $\iota = T$ and a dispute; if DSB quality is low in the sense that $qk(s) > \bar{\mu}^{T}_{2,Pro}(s)$, we have $\iota = FT$ and no dispute.*

Note that if $\delta = 0$ and there is hence no litigation delay, the above characterization of equilibrium behavior collapses to our earlier analysis of BITs under ISDS.[33] On the other hand, if δ is sufficiently close to 1, (5.9) and (5.10) together with our focus on the relatively-low-dispute-cost case imply that the host government will always choose T when it has the discretion to do so and the foreign investor will never invoke the DSB, and hence for δ in this range and conditional on any level of investment, the BIT with prospective damages would be

[33] This is easy to see once it is recalled that under Assumption $2^{/}$ we have $\gamma^*_{I^*} = 1$.

valueless, as it would deliver the noncooperative outcome in which the host government always expropriates in σ^T and expropriates in σ^{FT} with probability $1-\bar{p}$. Formally, the critical level of δ beyond which a BIT with prospective damages would be valueless, which we denote by $\bar{\delta}$, is defined by

$$\bar{\delta} = 1 - \min_{s} \left[c^* \epsilon^*(s) \right] > 0$$

where the inequality follows from our focus on the relatively-low-dispute-cost case. For $\delta \in [\bar{\delta}, 1]$, it follows from the above characterization of equilibrium behavior that in all states $s \in \sigma^{FT}$ with probability $1-\bar{p}$ the host government chooses T and the foreign investor will not file (because $qk(s) > 0 \geq (\bar{\mu}^{FT}_{2,Pro}(s))$ for all s), and in all states $s \in \sigma^T$ the host government chooses T and the foreign investor does not file (because $qk(s) < 1 \leq \bar{\mu}^{T}_{1,Pro}(s)$ for all s).

Turning to the ex-ante stage where the level of I^* is determined, if the investor is not subject to a taking for some fraction of the period, then for this fraction of the period the investor collects the market-clearing price $\tilde{P}(I^*)$ for the output from the investment (regardless of the state of the world) and therefore for this fraction of the period earns operating revenue $\pi(I^*, FT) = PS(I^*)$ implying a return on FDI of $\tilde{P}(I^*)$. On the other hand, if for this fraction of the period the investor is subject to a taking, then for this fraction of the period the investor gets nothing. Using this and the ex-post treatment of FDI implied under prospective damages by the equilibrium outcomes above as well as the fact that the host government is forced to implement the first-best policies with probability $\bar{p}$ and can act at its own discretion with probability $1-\bar{p}$, if the investor were to anticipate this treatment then this would imply an expected return on a level of FDI I^* from expected ex-post operating profits, which we denote by $E_s[\rho^*_{Pro}(I^*, s)]$, of

$$\begin{aligned}
E_s[\rho^*_{Pro}(I^*,s)] \;=\; & (1-\bar{p}) \sum_{s \epsilon \{\bar{\sigma}^{FT}_{Pro,1} \cup \bar{\sigma}^{T}_{Pro,3}\}} p(s) \tilde{P}(I^*) \\
& + (1-\bar{p}) \sum_{s \epsilon \bar{\sigma}^{FT}_{Pro,2}} p(s) \{ [1-\delta][1-qk(s)] \tilde{P}(I^*) - \frac{c^*(I^*,s)}{I^*} \} \\
& + (1-\bar{p}) \sum_{s \epsilon \bar{\sigma}^{T}_{Pro,2}} p(s) \{ [1-\delta] qk(s) \tilde{P}(I^*) - \frac{c^*(I^*,s)}{I^*} \} \\
& + \bar{p} p^{FT} \tilde{P}(I^*) \;+\; \bar{p} \sum_{s \in \left\{ \bar{\sigma}^{T}_{Pro,2}, \bar{\sigma}^{T}_{Pro,3} \right\}} p(s) \left[qk(s) \tilde{P}(I^*) - \frac{c^*(I^*,s)}{I^*} \right].
\end{aligned}$$

Here, $\bar{\sigma}^{FT}_{Pro,1}$ denotes the set of states for which FT is efficient, the importer government chooses

FT, and no complaint is filed (i.e., s such that $s \in \sigma^{FT}$ and $qk(s) < \bar{\mu}^{FT}_{1,Pro}(s)$), and $\bar{\sigma}^{FT}_{Pro,2}$ denotes the set of states for which FT is efficient, the importer government chooses T, and a complaint is filed (i.e., s such that $s \in \sigma^{FT}$ and $qk(s) \in [\bar{\mu}^{FT}_{1,Pro}(s), \bar{\mu}^{FT}_{2,Pro}(s)]$). Similarly, $\bar{\sigma}^{T}_{Pro,2}$ denotes the set of states for which T is efficient, the importer government chooses T, and a complaint is filed (i.e., s such that $s \in \sigma^{T}$ and $qk(s) \in [\bar{\mu}^{T}_{1,Pro}(s), \bar{\mu}^{T}_{2,Pro}(s)]$), while $\bar{\sigma}^{T}_{Pro,3}$ denotes the set of states for which T is efficient, the importer government chooses FT, and no complaint is filed (i.e., s such that $s \in \sigma^{T}$ and $qk(s) > \bar{\mu}^{T}_{2,Pro}(s)$).

Finally, denoting by $\omega_{Pro}(I^*, s)$ the host government payoff in state s conditional on a level of investment I^* in the presence of a BIT with prospective damages, and recalling that the investor must expect to receive a return of r^* once up-front payments and expected ex-post operating profits are accounted for, we can use the equilibrium actions as characterized above to express the expected level of this payoff:

$$
\begin{aligned}
E_s[\omega_{Pro}(I^*, s)] \;=\; & (1-\bar{p}) \sum_{s \in \bar{\sigma}^{FT}_{Pro,1}} p(s) CS(I^*) \\
& + (1-\bar{p}) \sum_{s \in \bar{\sigma}^{FT}_{Pro,2}} p(s) \{ CS(I^*) + [\delta + (1-\delta) qk(s)][\kappa PS(I^*)] - c(I^*, s) \} \\
& + (1-\bar{p}) \sum_{s \in \bar{\sigma}^{FT}_{Pro,3}} p(s) [CS(I^*) + \kappa PS(I^*)] \\
& + (1-\bar{p}) \sum_{s \in \bar{\sigma}^{T}_{Pro,2}} p(s) \{ [1-\delta] qk(s) [CS(I^*) - e(I^*, s)] - c(I^*, s) \} \\
& + (1-\bar{p}) \sum_{s \in \bar{\sigma}^{T}_{Pro,3}} p(s) [CS(I^*) - e(I^*, s)] + \bar{p} p^{FT} CS(I^*) \\
& + \bar{p} \sum_{s \in \{\bar{\sigma}^{T}_{Pro,2}, \bar{\sigma}^{T}_{Pro,3}\}} p(s) \{ qk(s) [CS(I^*) - e(I^*, s)] - c(I^*, s) \} \\
& - \{ r^* - E_s[\rho^*_{Pro}(I^*, s)] \} I^*.
\end{aligned}
$$

Here, the new set $\bar{\sigma}^{FT}_{Pro,3}$ denotes the set of states for which FT is efficient, the importer government chooses T, and no complaint is filed (i.e., s such that $s \in \sigma^{FT}$ and $qk(s) > \bar{\mu}^{FT}_{2,Pro}(s)$). And the term $\{r^* - E_s[\rho^*_{Pro}(I^*, s)]\} I^*$ reflects the up-front payment (if positive, tax if negative) required to attract the investment level I^*.

Plugging the expression for $E_s[\rho^*_{Pro}(I^*,s)]$ into $E_s[\omega_{Pro}(I^*,s)]$ yields

$$\begin{aligned}
E_s[\omega_{Pro}(I^*,s)] &= (1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Pro,1}} p(s)[CS(I^*)+PS(I^*)] \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Pro,2}} p(s)\{CS(I^*)+[1-(1-\kappa)[\delta+(1-\delta)qk(s)]]PS(I^*) \\
&-c^*(I^*,s)-c(I^*,s)\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Pro,3}} p(s)[CS(I^*)+\kappa PS(I^*)] \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{Pro,2}} p(s)\{[1-\delta]qk(s)[CS(I^*)+PS(I^*)-e(I^*,s)]-c^*(I^*,s)-c(I^*,s)\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{Pro,3}} p(s)[CS(I^*)+PS(I^*)-e(I^*,s)]+\bar{p}p^{FT}\left[CS\left(I^*\right)+PS\left(I^*\right)\right] \\
&+\bar{p}\sum_{s\in\left\{\bar{\sigma}^{T}_{Pro,2},\bar{\sigma}^{T}_{Pro,3}\right\}} p\left(s\right)\left\{qk\left(s\right)\left[CS\left(I^*\right)+PS\left(I^*\right)-e\left(I^*,s\right)\right]-c\left(I^*,s\right)-c^*\left(I^*,s\right)\right\} \\
&-r^*I^*.
\end{aligned} \tag{5.11}$$

Using the expression for $E_s[\omega_{Pro}(I^*,s)]$ in (5.11), we can now solve for the optimal level of FDI in the presence of a BIT with prospective damages, which we denote by $\bar{I}^*_{Pro}$, defined implicitly by $\frac{dE_s[\omega_{Pro}(I^*,s)]}{dI^*}=0$. The associated first order conditions can be manipulated to yield the following implicit characterization of $\bar{I}^*_{Pro}$:

$$\begin{aligned}
\tilde{P}(\bar{I}^*_{Pro}) &= \frac{r^*}{p^{FT}}+(1-\bar{p})(1-\kappa)\frac{[p^{FT}_{3,Pro}+\sum_{s\in\bar{\sigma}^{FT}_{2,Pro}}p(s)[\delta+(1-\delta)qk(s)]]}{p^{FT}}\frac{\partial PS(\bar{I}^*_{Pro})}{\partial I^*} \\
&+(1-\bar{p})\sum_{s\in\{\bar{\sigma}^{FT}_{2,Pro}\cup\bar{\sigma}^{T}_{2,Pro}\}}\frac{p(s)}{p^{FT}}[\frac{\partial c^*(\bar{I}^*_{Pro},s)}{\partial I^*}+\frac{\partial c(\bar{I}^*_{Pro},s)}{\partial I^*}] \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{2,Pro}}\frac{p(s)[1-\delta]qk(s)}{p^{FT}}[e(s)-\tilde{P}(\bar{I}^*_{Pro})] \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{3,Pro}}\frac{p(s)}{p^{FT}}[e(s)-\tilde{P}(\bar{I}^*_{Pro})] \\
&+\bar{p}\sum_{s\in\left\{\bar{\sigma}^{T}_{2,Pro},\bar{\sigma}^{T}_{3,Pro}\right\}}\frac{p\left(s\right)}{p^{FT}}\left[\frac{\partial c^*\left(\bar{I}^*_{Pro},s\right)}{\partial I^*}+\frac{\partial c\left(\bar{I}^*_{Pro},s\right)}{\partial I^*}\right] \\
&+\bar{p}\sum_{s\in\left\{\bar{\sigma}^{T}_{2,Pro},\bar{\sigma}^{T}_{3,Pro}\right\}}\frac{p\left(s\right)qk\left(s\right)}{p^{FT}}\left[e\left(s\right)-\tilde{P}\left(\bar{I}^*_{Pro}\right)\right].
\end{aligned} \tag{5.12}$$

Finally, plugging $\bar{I}^*_{Pro}$ into the expression for $E_s[\omega_{Pro}(I^*,s)]$ in (5.11) and simplifying yields

$$\begin{aligned} E_s[\omega_{Pro}(\bar{I}^*_{Pro},s)] &= (1-\bar{p})[p^{FT}+p^T_{3,Pro}+\sum_{s\in\bar{\sigma}^T_{2,Pro}} p(s)[1-\delta]qk(s)]CS(\bar{I}^*_{Pro}) \\ &+(1-\bar{p})\left[p^{FT}_{3,Pro}+\sum_{s\in\bar{\sigma}^{FT}_{2,Pro}} p(s)[\delta+(1-\delta)qk(s)]\right](1-\kappa)(\bar{I}^*_{Pro})^2\frac{\partial\tilde{P}(\bar{I}^*_{Pro})}{\partial I^*} \\ &-(1-\bar{p})\sum_{s\in\{\bar{\sigma}^{FT}_{2,Pro}\cup\bar{\sigma}^T_{2,Pro}\}} p(s)\left[c^*(\bar{I}^*_{Pro},s)+c(\bar{I}^*_{Pro},s)-\bar{I}^*_{Pro}[\frac{\partial c^*(\bar{I}^*_{Pro},s)}{\partial I^*}+\frac{\partial c(\bar{I}^*_{Pro},s)}{\partial I^*}]\right] \\ &+\bar{p}\left[p^{FT}+\sum_{s\in\left\{\bar{\sigma}^T_{2,Pro},\bar{\sigma}^T_{3,Pro}\right\}} p\left(s\right)qk\left(s\right)\right]CS\left(I^*\right) \\ &-\bar{p}\sum_{s\in\left\{\bar{\sigma}^T_{2,Pro},\bar{\sigma}^T_{3,Pro}\right\}} p\left(s\right)\left\{c\left(I^*,s\right)+c^*\left(I^*,s\right)-\bar{I}^*_{Pro}\left[\frac{\partial c^*\left(\bar{I}^*_{Pro},s\right)}{\partial I^*}+\frac{\partial c\left(\bar{I}^*_{Pro},s\right)}{\partial I^*}\right]\right\}. \end{aligned} \tag{5.13}$$

Notice from (5.12) that for any level of court quality $q\in(0,1)$, if the degree of litigation delay is sufficiently high so that $\delta\in[\bar{\delta},1]$, we then have

$$\tilde{P}(\bar{I}^*_{Pro})=\frac{r^*}{p^{FT}}+(1-\bar{p})(1-\kappa)\frac{\partial PS(\bar{I}^*_{Pro})}{\partial I^*}$$

which by (3.7) implies $\bar{I}^*_{Pro}=\hat{I}^*$; and from (5.13) we have

$$E_s[\omega_{Pro}(\bar{I}^*_{Pro},s)]=p^{FT}[CS(\bar{I}^*_{Pro})+(1-\bar{p})(1-\kappa)(\bar{I}^*_{Pro})^2\frac{\partial\tilde{P}(\bar{I}^*_{Pro})}{\partial I^*}]$$

which by (3.8) implies $E_s[\omega_{Pro}(\bar{I}^*_{Pro},s)]=E_s[\omega(\hat{I}^*,T,s)]$. Clearly, then, if litigation delay is sufficiently high, the host government cannot improve upon a program of up-front investment incentives to foreign investors by introducing a BIT with prospective damages, no matter how accurate the court may be.

We record this in:

Remark 1. *If litigation delay is sufficiently high (for $\delta\geq\bar{\delta}$), the host government cannot improve upon a program of up-front investment incentives to foreign investors by introducing a BIT with prospective damages, no matter how accurate the court may be (for $q>0$).*

Retrospective damages We consider next the case of retrospective damages. Under retrospective damages, if the DSB rules for FT, the host government must both cease and desist its T policy and revert to FT henceforth (prospective damages), *and* it must make damage payments to the foreign investor in the amount of the harm $\delta|\gamma^*_{I^*}(I^*)|$ already suffered (retrospective damages), costing the host government $\frac{1}{\beta}[\delta|\gamma^*_{I^*}(I^*)|]$ in lost surplus. Consistent with the standard practice in BITs discussed above and with our findings of section 4, we assume that in the context of an investment dispute these damage payments take the form of cash. Hence, while we continue to characterize model outcomes for $\beta \in (0, 1]$, we will highlight results for the case where β is large, reflecting the relative efficiency of cash transfers as compared to the tariff adjustments typically used for such transfers in the context of trade disputes.

Adopting retrospective damages will have implications for the equilibrium behavior of the host government and the foreign investor and for the payoffs to the host government and to the foreign investor in any state conditional on this equilibrium behavior. Again we consider each in turn.

Consider first the foreign investor's filing behavior under retrospective damages. The foreign investor files a complaint if and only if $\iota = T$ and its expected benefit of filing exceeds its cost of filing, that is

$$\Pr(\text{DSB ruling is } FT \mid s) \times |\gamma^*_{I^*}(I^*)| > c^*(I^*, s). \tag{5.14}$$

Condition (5.14) is the "filing" condition for the foreign investor to invoke the DSB in response to a policy choice by the host government of $\iota = T$.

Next consider the host government's policy choice under retrospective damages. Whenever it has discretion to do so, the host government chooses $\iota = T$ if either (5.1) fails – because then the host government can set $\iota = T$ without triggering a dispute – or if (5.1) holds and the expected benefit to the host government from a taking exceeds the cost to the host government of a dispute:

$$\delta\{[\gamma_G(I^*, s) - \Pr(\text{DSB ruling is } FT \mid s) \times \frac{|\gamma^*_{I^*}(I^*)|}{\beta}\} + [1-\delta]\Pr(\text{DSB ruling is } T \mid s) \times \gamma_G(I^*, s) > c(I^*, s)$$

or

$$\Pr(\text{DSB ruling is } T \mid s) \times \gamma_G(I^*, s) + \Pr(\text{DSB ruling is } FT \mid s) \times \{\delta[\gamma_G(I^*, s) - \frac{|\gamma^*_{I^*}(I^*)|}{\beta}]\} > c(I^*, s).$$

We can now derive the equilibrium actions, conditional on investment level I^*, for each state s. Again we assume that the states where the vague contract is unambiguous are measure zero, so we can focus only on states where the court if invoked must interpret the contract. Recalling that the host government is constrained to implement the first-best policies with probability $\bar{p}$ and can act at its own discretion with probability $1-\bar{p}$, defining the thresholds , $\bar{\mu}^{FT}_{1,Retro}(s) \equiv \frac{\delta+[c\epsilon(s)-\delta]\kappa\beta}{\delta+[1-\delta]\kappa\beta}$, $\bar{\mu}^{FT}_{2,Retro}(s) \equiv 1-c^*\epsilon^*(s)$, $\bar{\mu}^{T}_{1,Retro}(s) \equiv c^*\epsilon^*(s)$, and $\bar{\mu}^{T}_{2,Retro}(s) \equiv \frac{1-c\epsilon(s)}{[1-\delta]+\frac{\delta}{\beta}\frac{PS(I^*)}{e(I^*,s)-CS(I^*)}}$, and noting as before that $\bar{\mu}^{FT}_{1,Retro}(s) < \bar{\mu}^{FT}_{2,Retro}(s)$ and $\bar{\mu}^{T}_{1,Retro}(s) < \bar{\mu}^{T}_{2,Retro}(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, the conditions above immediately imply the following result:

Lemma 9. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\bar{\mu}^{FT}_{1,Retro}(s) < \bar{\mu}^{FT}_{2,Retro}(s)$ and $\bar{\mu}^{T}_{1,Retro}(s) < \bar{\mu}^{T}_{2,Retro}(s)$, equilibrium actions can be characterized as follows:*

1. *In states $s \in \sigma^{FT}$:*
    1. *If the host government is constrained: We have $\iota = FT$ and no dispute.*
    2. *If the host government is unconstrained: If DSB quality is high in the sense that $qk(s) < \bar{\mu}^{FT}_{1,Retro}(s)$, we have $\iota = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\bar{\mu}^{FT}_{1,Retro}(s), \bar{\mu}^{FT}_{2,Retro}(s)\right]$, we have $\iota = T$ and a dispute; if DSB quality is low in the sense that $qk(s) > \bar{\mu}^{FT}_{2,Retro}(s)$, we have $\iota = T$ and no dispute.*
2. *In states $s \in \sigma^{T}$:*
    1. *If the host government is constrained: We have $\iota = T$, no dispute if $qk(s) < \bar{\mu}^{T}_{1,Retro}(s)$, and a dispute if $qk(s) > \bar{\mu}^{T}_{1,Retro}(s)$.*
    2. *If the host government is unconstrained: If DSB quality is high in the sense that $qk(s) < \bar{\mu}^{T}_{1,Retro}(s)$, we have $\iota = T$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\bar{\mu}^{T}_{1,Retro}(s), \bar{\mu}^{T}_{2,Retro}(s)\right]$, we have $\iota = T$ and a dispute; if DSB quality is low in the sense that $qk(s) > \bar{\mu}^{T}_{2,Retro}(s)$, we have $\iota = FT$ and no dispute.*

Note that as β approaches 0, $\bar{\mu}^{FT}_{1,Retro}(s)$ approaches 1 and $\bar{\mu}^{T}_{2,Retro}(s)$ approaches 0, and in this case regardless of how good the court may be, the host government always chooses FT to avoid the possibility of costly payments to the foreign investor for retroactive damages.[34]

Turning to the ex-ante stage where the level of I^* is determined and using the ex-post treatment of FDI implied under retrospective damages by the equilibrium outcomes above as well as the fact that the host government is forced to implement the first-best policies with probability $\bar{p}$ and can act at its own discretion with probability $1-\bar{p}$, if the investor were to anticipate this treatment then this would imply an expected return on a level of FDI I^* from expected ex-post operating profits, which we denote by $E_s[\rho^*_{Retro}(I^*,s)]$, of

$$\begin{aligned} E_s[\rho^*_{Retro}(I^*,s)] &= (1-\bar{p}) \sum_{s\epsilon\{\bar{\sigma}^{FT}_{Retro,1}\cup\bar{\sigma}^{T}_{Retro,3}\}} p(s)\tilde{P}(I^*) \\ &\quad +(1-\bar{p}) \sum_{s\epsilon\bar{\sigma}^{FT}_{Retro,2}} p(s)\{[1-qk(s)]\tilde{P}(I^*) - \frac{c^*(I^*,s)}{I^*}\} \\ &\quad +(1-\bar{p}) \sum_{s\epsilon\bar{\sigma}^{T}_{Retro,2}} p(s)\{qk(s)\tilde{P}(I^*) - \frac{c^*(I^*,s)}{I^*}\} \\ &\quad +\bar{p}p^{FT}\tilde{P}(I^*) \quad + \quad \bar{p} \sum_{s\in\left\{\bar{\sigma}^{T}_{Retro,2},\bar{\sigma}^{T}_{Retro,3}\right\}} p(s)\left[qk(s)\tilde{P}(I^*) - \frac{c^*(I^*,s)}{I^*}\right]. \end{aligned}$$

Here, $\bar{\sigma}^{FT}_{Retro,1}$ denotes the set of states for which FT is efficient, the importer government chooses FT, and no complaint is filed (i.e., s such that $s\in\sigma^{FT}$ and $qk(s)<\bar{\mu}^{FT}_{1,Retro}(s)$, and $\bar{\sigma}^{FT}_{Retro,2}$ denotes the set of states for which FT is efficient, the importer government chooses T, and a complaint is filed (i.e., s such that $s\in\sigma^{FT}$ and $qk(s)\in[\bar{\mu}^{FT}_{1,Retro}(s),\bar{\mu}^{FT}_{2,Retro}(s)]$). Similarly, $\bar{\sigma}^{T}_{Retro,2}$ denotes the set of states for which T is efficient, the importer government chooses T, and a complaint is filed (i.e., s such that $s\in\sigma^{T}$ and $qk(s)\in[\bar{\mu}^{T}_{1,Retro}(s),\bar{\mu}^{T}_{2,Retro}(s)]$), while $\bar{\sigma}^{T}_{Retro,3}$ denotes the set of states for which T is efficient, the importer government chooses FT, and no complaint is filed (i.e., s such that $s\in\sigma^{T}$ and $qk(s)>\bar{\mu}^{T}_{2,Retro}(s)$).

Finally, denoting by $\omega_{Retro}(I^*,s)$ the host government payoff in state s conditional on a level of investment I^* in the presence of a BIT with retrospective damages, and recalling that the investor must expect to receive a return of r^* once up-front payments and expected ex-post operating profits are accounted for, we can use the equilibrium actions as characterized above

[34]Notice too that as we allow β to approach 0 we must also have $c^*(s)$ approaching 0 in order to maintain our "small litigation cost" focus and ensure that $\bar{\mu}^{FT}_{1,Retro}(s)<\bar{\mu}^{FT}_{2,Retro}(s)$ and $\bar{\mu}^{T}_{1,Retro}(s)<\bar{\mu}^{T}_{2,Retro}(s)$.

to express the expected level of this payoff:

$$\begin{aligned}
E_s[\omega_{Retro}(I^*,s)] &= (1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Retro,1}} p(s)CS(I^*) + (1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Retro,2}} p(s)\{CS(I^*) \\
&+\left[qk(s)\kappa + [1-qk(s)]\delta[\kappa - \frac{1}{\beta}]\right] PS(I^*) - c(I^*,s)\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Retro,3}} p(s)\{CS(I^*) + \kappa PS(I^*)\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{Retro,2}} p(s)\{qk(s)\left[[1-\delta][CS(I^*) - e(I^*,s)] - \frac{\delta}{\beta}PS(I^*)\right] - c(I^*,s)\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{Retro,3}} p(s)\{CS(I^*) - e(I^*,s)\} + \bar{p}p^{FT}CS(I^*) \\
&+\bar{p}\sum_{s\in\{\bar{\sigma}^{T}_{Retro,2},\bar{\sigma}^{T}_{Retro,3}\}} p(s)\{qk(s)[CS(I^*) - e(I^*,s)] - c(I^*,s)\} \\
&-\{r^* - E_s[\rho^*_{Retro}(I^*,s)]\}I^*.
\end{aligned}$$

Here, the new set $\bar{\sigma}^{FT}_{Retro,3}$ denotes the set of states s such that $s \in \sigma^{FT}$ and $qk(s) > \bar{\mu}^{FT}_{2,Retro}(s)$.

Plugging the expression for $E_s[\rho^*_{Retro}(I^*,s)]$ into the expression for $E_s[\omega_{Retro}(I^*,s)]$ yields

$$\begin{aligned}
E_s[\omega_{Retro}(I^*,s)] &= (1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Retro,1}} p(s)[CS(I^*) + PS(I^*)] \qquad (5.15)\\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Retro,2}} p(s)\{CS(I^*) + \left[qk(s)\kappa + [1-qk(s)][1+\delta(\kappa - \frac{1}{\beta})\right] PS(I^*) \\
&-c^*(I^*,s) - c(I^*,s)\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{Retro,3}} p(s)\{CS(I^*) + \kappa PS(I^*)\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{Retro,2}} p(s)\{qk(s)\left[[1-\delta][CS(I^*) - e(I^*,s)] + (1-\frac{\delta}{\beta})PS(I^*)\right] \\
&-c^*(I^*,s) - c(I^*,s)\} \\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{Retro,3}} p(s)[CS(I^*) + PS(I^*) - e(I^*,s)] \\
&\bar{p}\sum_{s\in\left\{\bar{\sigma}^{T}_{Retro,2},\bar{\sigma}^{T}_{Retro,3}\right\}} p\left(s\right)\left\{qk\left(s\right)\left[CS\left(I^*\right) + PS\left(I^*\right) - e\left(I^*,s\right)\right] - c\left(I^*,s\right) - c^*\left(I^*,s\right)\right\} \\
&-r^*I^*.
\end{aligned}$$

Using (5.15), we can now solve for $\bar{I}^*_{Retro}$, the optimal level of FDI in the presence of a BIT with

retrospective damages, defined implicitly by $\frac{dE_s[\omega_{Retro}(I^*,s)]}{dI^*}=0$ and yielding

$$\begin{aligned}
\tilde{P}(\bar{I}^*_{Retro}) &= \frac{r^*}{p^{FT}}+(1-\bar{p})\sum_{s\in\bar{\sigma}^T_{3,Retro}}\frac{p(s)}{p^{FT}}[e(s)-\tilde{P}(\bar{I}^*_{Retro})] \qquad (5.16)\\
&+(1-\bar{p})\{\frac{(1-\kappa)p^{FT}_{3,Retro}+\sum_{s\in\bar{\sigma}^{FT}_{2,Retro}}p(s)[qk(s)(1-\kappa)+\delta(1-qk(s))(\frac{1}{\beta}-\kappa)}{p^{FT}}\\
&+\frac{\sum_{s\in\bar{\sigma}^T_{2,Retro}}p(s)qk(s)\delta(\frac{1}{\beta}-1)}{p^{FT}}\}\frac{\partial PS(\bar{I}^*_{Retro})}{\partial I^*}\\
&+(1-\bar{p})\sum_{s\in\{\bar{\sigma}^{FT}_{2,Retro}\cup\bar{\sigma}^T_{2,Retro}\}}\frac{p(s)}{p^{FT}}[\frac{\partial c^*(\bar{I}^*_{Retro},s)}{\partial I^*}+\frac{\partial c(\bar{I}^*_{Retro},s)}{\partial I^*}]\\
&+(1-\bar{p})\sum_{s\in\bar{\sigma}^T_{2,Retro}}\frac{p(s)[1-\delta]qk(s)}{p^{FT}}[e(s)-\tilde{P}(\bar{I}^*_{Retro})]\\
&+\bar{p}\sum_{s\in\left\{\bar{\sigma}^T_{2,Retro},\bar{\sigma}^T_{3,Retro}\right\}}\frac{p\left(s\right)}{p^{FT}}\left[\frac{\partial c^*\left(\bar{I}^*_{Retro},s\right)}{\partial I^*}+\frac{\partial c\left(\bar{I}^*_{Retro},s\right)}{\partial I^*}\right]\\
&+\bar{p}\sum_{s\in\left\{\bar{\sigma}^T_{2,Retro},\bar{\sigma}^T_{3,Retro}\right\}}\frac{p\left(s\right)qk\left(s\right)}{p^{FT}}\left[e\left(s\right)-\tilde{P}\left(\bar{I}^*_{Retro}\right)\right].
\end{aligned}$$

Finally, plugging $\bar{I}^*_{Retro}$ into the expression for $E_s[\omega_{Retro}(I^*,s)]$ in (5.15) and simplifying yields

$E_s[\omega_{Retro}(\bar{I}^*_{Retro},s)]=$

$$\begin{aligned}
&(1-\bar{p})\{p^{FT}+p^T_{3,Retro}+\sum_{s\in\bar{\sigma}^T_{2,Retro}}p(s)[1-\delta]qk(s)\}CS(\bar{I}^*_{Retro}) \qquad (5.17)\\
&+(1-\bar{p})\{\sum_{s\in\bar{\sigma}^{FT}_{2,Retro}}p(s)[qk(s)(1-\kappa)+(1-qk(s))\delta(\frac{1}{\beta}-\kappa)]\\
&+\sum_{s\in\bar{\sigma}^T_{2,Retro}}p(s)qk(s)\delta(\frac{1}{\beta}-1)-p^{FT}_{3,Retro}(1-\kappa)\}(\bar{I}^*_{Retro})^2\frac{\partial\tilde{P}(\bar{I}^*_{Retro})}{\partial I^*}\\
&-(1-\bar{p})\sum_{s\in\{\bar{\sigma}^{FT}_{2,Retro}\cup\bar{\sigma}^T_{2,Retro}\}}p(s)\{[c^*(\bar{I}^*_{Retro},s)+c(\bar{I}^*_{Retro},s)]-\bar{I}^*_{Retro}[\frac{\partial c^*(\bar{I}^*_{Retro},s)}{\partial I^*}+\frac{\partial c(\bar{I}^*_{Retro},s)}{\partial I^*}]\}\\
&+\bar{p}\left[p^{FT}+\sum_{s\in\left\{\bar{\sigma}^T_{2,Retro},\bar{\sigma}^T_{3,Retro}\right\}}p\left(s\right)qk\left(s\right)\right]CS\left(I^*\right)\\
&-\bar{p}\sum_{s\in\left\{\bar{\sigma}^T_{2,Retro},\bar{\sigma}^T_{3,Retro}\right\}}p\left(s\right)\left\{c\left(\bar{I}^*_{Retro},s\right)+c^*\left(\bar{I}^*_{Retro},s\right)-\bar{I}^*_{Retro}\left[\frac{\partial c^*\left(\bar{I}^*_{Retro},s\right)}{\partial I^*}+\frac{\partial c\left(\bar{I}^*_{Retro},s\right)}{\partial I^*}\right]\right\}.
\end{aligned}$$

Recall now that as β approaches 0, the host government will always choose FT to avoid any possibility of having to make costly transfer payments to the foreign investor for retroactive damages, regardless of how good the court may be. In particular, as $\beta \rightarrow 0$ we have $\bar{\sigma}^{FT}_{1,Retro} \rightarrow \sigma^{FT}$ and $\bar{\sigma}^{T}_{3,Retro} \rightarrow \sigma^{T}$, which by (5.16) and (5.17) implies that $\bar{I}^{*}_{Retro}$ is then implicitly defined by

$$\begin{aligned} \tilde{P}(\bar{I}^{*}_{Retro}) &= \frac{r^{*}}{p^{FT}} + \sum_{s \in \sigma^{T}} \frac{p(s)}{p^{FT}}[e(s) - \tilde{P}(\bar{I}^{*}_{Retro})] \qquad (5.18) \\ &\quad -\bar{p} \sum_{s \in \sigma^{T}} \frac{p(s)}{p^{FT}} \{[1 - qk(s)][e(s) - \tilde{P}(\bar{I}^{*}_{Retro})] - [\frac{\partial c^{*}\left(\bar{I}^{*}_{Retro}, s\right)}{\partial I^{*}} + \frac{\partial c\left(\bar{I}^{*}_{Retro}, s\right)}{\partial I^{*}}]\} \end{aligned}$$

and $E_{s}[\omega_{Retro}(\bar{I}^{*}_{Retro}, s)]$ is given by

$$\begin{aligned} E_{s}[\omega_{Retro}(\bar{I}^{*}_{Retro}, s)] &= [1 - \bar{p} \sum_{s \in \sigma^{T}} p(s)(1 - qk(s))] CS(\bar{I}^{*}_{Retro}) \qquad (5.19) \\ &\quad -\bar{p} \sum_{s \in \sigma^{T}} p(s) \{c\left(\bar{I}^{*}_{Retro}, s\right) + c^{*}\left(\bar{I}^{*}_{Retro}, s\right) - \bar{I}^{*}_{Retro}[\frac{\partial c^{*}\left(\bar{I}^{*}_{Retro}, s\right)}{\partial I^{*}} + \frac{\partial c\left(\bar{I}^{*}_{Retro}, s\right)}{\partial I^{*}}]\}. \end{aligned}$$

If the possibility of sufficiently damaging externality realizations exist so that $e(s)$ is very large in some states, then (5.18) implies that $\bar{I}^{*}_{Retro}$ will be driven toward zero and (5.19) then implies that the expected surplus of the host government under retrospective damages will be driven toward zero as well.[35] We record this in:

Remark 2. *If the cost of transfers is sufficiently high (for β approaching 0), a BIT with retrospective damages induces the host government to avoid takings in any state, no matter how accurate the court may be (for $q > 0$); and if there is a possibility of sufficiently large damaging externality realizations in some states (if $e(s)$ is sufficiently large in some states), then the expected surplus of the host government under retrospective damages will be driven to zero.*

On the other hand, fixing a level of $\beta > 0$ and for any level of δ including $\delta \in [\bar{\delta}, 1]$, if the court quality is high enough ($q \rightarrow 0$) then $\sigma^{FT}_{1} \rightarrow \sigma^{FT}$ and $\sigma^{T}_{1} \rightarrow \sigma^{T}$, which by (5.16) and

[35] Notice the absence of the externality variable $e(s)$ entering directly into the expression for host government expected welfare in (5.19), despite the fact that for the case under consideration with β approaching 0 the host government always chooses FT to avoid any possibility of having to make costly transfer payments to the foreign investor for retroactive damages and hence experiences the negative consequences of the externality for $s \in \sigma^{T}$. In fact, as (5.19) indicates when viewed in the context of (5.18), there is a cost to the host government associated with these externalities, but it is a cost which takes the form of a lower optimally chosen investment level $\bar{I}^{*}_{Retro}$ and the lower home country consumer surplus that accompanies this lower investment.

(5.17) then implies that the first best investment given by (**??**) and first best host government welfare given by (**??**) can be achieved under retrospective damages. We record this in:

Remark 3. *Fixing a level of $\beta > 0$ and for any level of δ including $\delta \in [\bar{\delta}, 1]$, if the quality of the court is sufficiently high ($q \to 0$), then $\sigma_1^{FT} \to \sigma^{FT}$ and $\sigma_1^{T} \to \sigma^{T}$, and the expected surplus of the host government under retrospective damages will reach the first best level.*

Using (5.13) and (5.17) together with Remarks 1, 2 and 3, we now may state the following conditions under which retrospective damages will be optimal for a BIT:

Proposition 7. *A retrospective remedy is optimal for a BIT ($E_s[\omega_{Retro}(\bar{I}^*_{Retro}, s)] > E_s[\omega_{Pro}(\bar{I}^*_{Pro}, s)]$) provided that the degree of litigation delay is sufficiently long (δ large), transfers in the context of a BIT are sufficiently efficient (β large) and the quality of the court is sufficiently high (q low).*

Intuitively, for any fixed court quality q, as δ approaches $\bar{\delta}$ and litigation delay becomes sufficiently long, the expected payoff to the host government under a BIT with ISDS and prospective damages will approach the expected payoff it would receive from simply offering a program of up-front investment incentives to foreign investors with no BIT. And provided that β is not too small and hence the cost of transfers is not too high, if the quality of the court is fixed at a sufficiently high level and for any level of $\delta \in [0, 1]$, the expected payoff to the host government under a BIT with ISDS and retrospective damages can be brought arbitrarily close to the first best level, which exceeds the expected payoff under a program of up-front investment incentives to foreign investors with no BIT and therefore beats prospective damages.

Summarizing, Propositions 6 and 7 imply the following: If court quality is sufficiently high (q is sufficiently low), prospective damages are employed in trade agreements such as the WTO while retrospective damages are employed in BITs because (i) the degree of pre-ruling harm is typically more severe in the context of investment disputes as compared to trade disputes (δ is relatively high for investment disputes but low for trade disputes) so that prospective damages become a poor option for effective investment agreements, and (ii) the available means of making international transfers are much less efficient in the context of trade disputes as compared to investment disputes (β is relatively low for trade disputes but high for investment disputes), making retrospective damages prohibitively expensive in the context of trade agreements and therefore unattractive in that context.

6. Conclusion

International investment agreements employ dispute settlement procedures that differ markedly from their counterparts in trade agreements along three key dimensions: standing, the nature of the remedy, and the remedial period. In this paper we have developed parallel models of trade agreements and investment agreements and have employed them to study these differences. We have argued that the differences can be understood as arising from the fundamentally different problems that trade and investment agreements are designed to solve.

Our results suggest that there are plausible conditions under which the broad differences between the dispute settlement procedures of trade and investment agreements can be viewed as an optimal response to the different environments within which these agreements operate. We find that this is the case when the quality of the court is generally high, and provided that courts have greater difficulty in assessing the monetary value of the harm suffered by a government from trade protection than they do in assessing the monetary value of the harm suffered by investors from a taking, that the proportion of the harm incurred prior to a court ruling tends to be large in the case of a taking relative to the case of trade protection, that free trade is likely to be the efficient trade policy outcome while the probability that a taking is the efficient policy outcome is low but non-trivial, and that the host government is highly inefficient in orchestrating takings when takings are not socially beneficial.

In this way, our results can help interpret the design differences across dispute settlement procedures in trade and investment agreements and provide support for the position that these differences are not arbitrary. At the same time, our results indicate that some of the most controversial features of these procedures, such as providing standing for investors to bring claims against foreign governments in investment disputes, are far from universally optimal under all circumstances. This suggests that such features deserve closer scrutiny before drawing the conclusion that they are warranted in the circumstances where they are present.

References

Bagwell, Kyle and Robert W. Staiger (2010). "Backward Stealing and Forward Manipulation in the WTO," *Journal of International Economics*, September.

Bond, Eric W. (1986). "Tax Holidays as Signals," *American Economic Review* 76(4), September.

Brainard, Lael S. (1997). "An Empirical Assessment of the Proximity-Concentration Trade-off Between Multinational Sales and Trade," *American Economic Review* 87(4), September.

Davey, William J. (2010). "Sanctions in the WTO: Problems and Solutions," in Chad Bown & Joost Pauwelyn eds. *The Law, Economics and Politics of Retaliation in WTO Dispute Settlement*: Cambridge Press.

Helpman, Elhanan, Melitz, Marc J. and Stephen R. Yeaple (2004). "Export Versus FDI with Heterogeneous Firms," *American Economic Review* 94(1), March.

Horn, Henrik and Thomas Tangeras (2016). "Economics and Politics of International Investment Agreements," Research Institute of Industrial Economics, IFN Working Paper No. 1140.

Janeba, Eckhard, (2016). "Regulatory Chill and the Effects of Investor State Dispute Settlements," CESifo Working Paper Series No. 6188, December.

Kohler, Wilhelm and Frank Stahler (2016). "The Economics of Investor Protection: ISDS versus National Treatment," CESifo Working Paper Series No. 5766, February.

Konrad, Kai A. (2017). "Large investors, regulatory taking and investor-state dispute settlement," *European Economic Review* 98.

Limao, Nuno and Kamal Saggi (2008). "Tariff retaliation versus financial compensation in the enforcement of international trade agreements," Journal of International Economics 76.

Maggi, Giovanni and Robert W. Staiger (2011). "The Role of Dispute Settlement Procedures in International Trade Agreements," *Quarterly Journal of Economics*, February.

Maggi, Giovanni and Robert W. Staiger (2018). "Trade Disputes and Settlement," *International Economic Review*, February.

McLaren, John (1997). "Size, Sunk Costs, and Judge Bowker's Objection to Free Trade," *American Economic Review* 87(3), June.

Ossa, Ralph, Staiger, Robert W. and Alan O. Sykes (2019). "FDI, Exports and the Terms-of-Trade Motive for Tariff Intervention," mimeo, March.

Park, Jee-Hyeong (2016). "Enforcement and Dispute Settlement," Chapter 1 in Kyle Bagwell and Robert W. Staiger (eds.) *Handbook of Commercial Policy* Vol 1B, Elsevier North Holland.

Schwartz, Warren F. and Alan O. Sykes (2002). "The Economic Structure of Renegotiation and Dispute Resolution in the World Trade Organization," *Journal of Legal Studies* 31.

Stahler, Frank (2018). "An optimal investor state dispute settlement mechanism," mimeo, March.

Staiger, Robert W. and Alan O. Sykes (2017). "How Important can the Non-Violation Clause be for the GATT/WTO," *American Economic Journal: Microeconomics*, May.

Sykes, Alan O. (2005). "Public vs. Private Enforcement of International Economic Law: Of Standing and Remedy," U Chicago Law & Economics, Olin Working Paper, No. 235.

UNCTAD (2018). IIA Mapping Project, available at https://investmentpolicyhub.unctad.org/IIA/mappedContent.

Yarbrough, Beth V., and Robert M. Yarbrough (1992). *Cooperation and Governance in International Trade: The Strategic Organizational Approach.* Princeton, NJ: Princeton University Press.

7. Figures

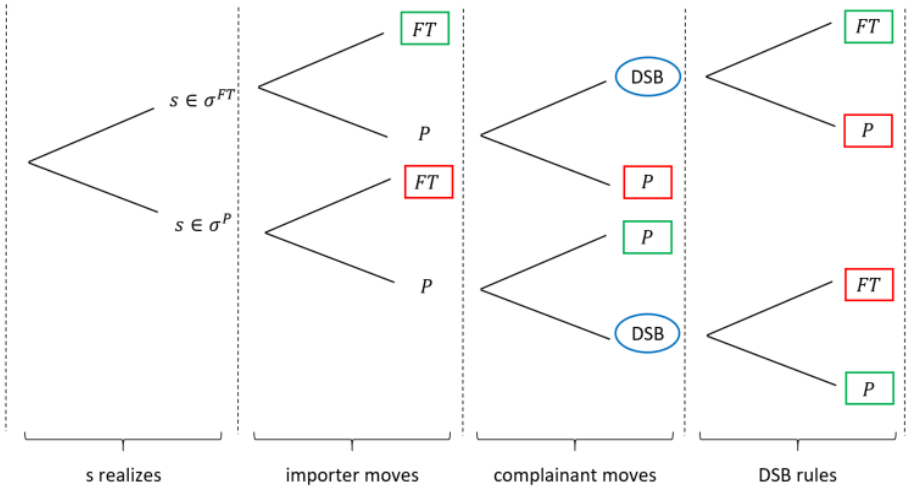


Figure 1

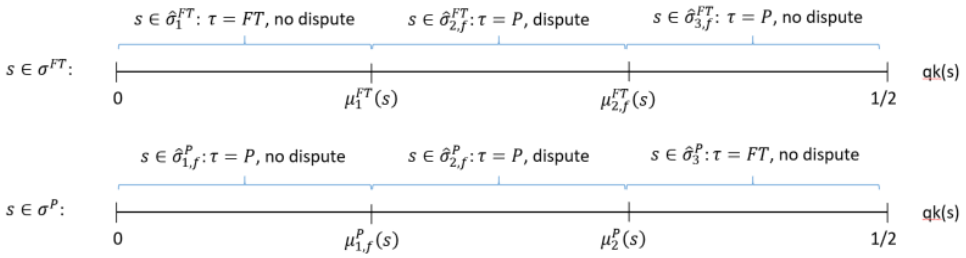


Figure 2a

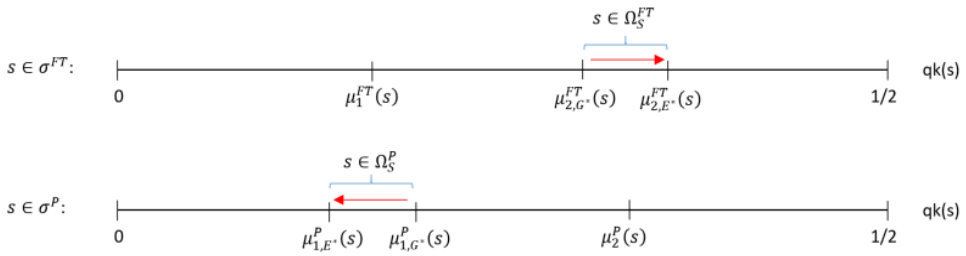


Figure 2b

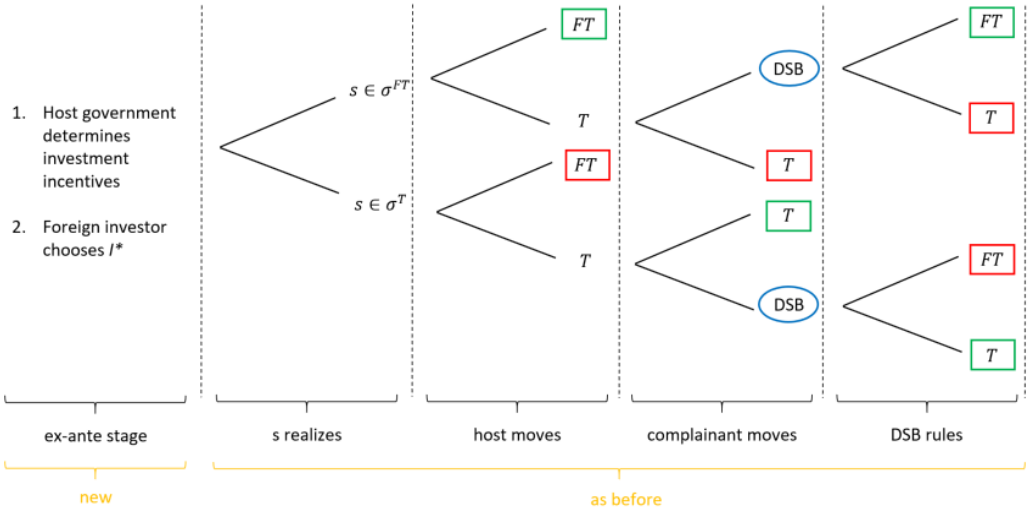

1. Host government determines investment incentives
2. Foreign investor chooses I*
$s \in \sigma^{FT}$
$s \in \sigma^{T}$
FT
T
FT
T
DSB
T
T
DSB
FT
T
FT
T
ex-ante stage
s realizes
host moves
complainant moves
DSB rules
new
as before

Figure 3