NBER WORKING PAPER SERIES

DISPUTES IN INTERNATIONAL INVESTMENT AND TRADE

Ralph Ossa
Robert W. Staiger
Alan O. Sykes

Working Paper 27012
http://www.nber.org/papers/w27012

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2020

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 views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

At least one co-author has disclosed a financial relationship of potential relevance for this research. Further information is available online at http://www.nber.org/papers/w27012.ack

Disputes in International Investment and Trade
Ralph Ossa, Robert W. Staiger, and Alan O. Sykes
NBER Working Paper No. 27012
April 2020
JEL No. F02,F1,F23

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.

Ralph Ossa
University of Zurich
Department of Economics
Schoenberggasse 1
8001 Zuerich
Switzerland
ralph.ossa@econ.uzh.ch

Robert W. Staiger
Department of Economics
Dartmouth College
319 Silsby
Hanover, NH 03755
and NBER
rstaiger@dartmouth.edu

Alan O. Sykes
Law School
Stanford University
Stanford, CA 94305
asykes@stanford.edu

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 such as those of the World Trade Organization (WTO), which secure their access to foreign markets. If they engage in foreign direct investment, they are protected by investment agreements such as those negotiated pursuant to the U.S. bilateral investment treaty (BIT) program and similar programs in the EU and elsewhere, 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 U.S.-Mexico-Canada Agreement (USMCA), and the Comprehensive and Progressive Agreement for Trans-Pacific Partnership (CPTPP), 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, 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 the USMCA, 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 USMCA, although it preserves

[1] In addition to the agreement establishing the 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.

the existing ISDS mechanism for claims involving Mexico only in a few sectors such as telecommunications and oil 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 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 on U.S. countervailing duty laws as "the latest example of judicial activism at the WTO seeking to undermine those laws."[8] Dissatisfaction with WTO dispute rulings has led the United States to block the appointment of judges to fill vacancies on the WTO Appellate Body for the 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 private rights of action in the form of ISDS rather than simply SSDS? Why do trade agreements uniformly provide for SSDS and not private rights of action in the form of exporter-state dispute settlement (ESDS)?[9] Are the critics of ISDS right that

[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 USMCA, for example, the investor-rights provisions allow for both SSDS and private rights of action in the form of ISDS. But the "trade"

it unwisely burdens national 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 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 in each setting of the features of dispute settlement that are at the heart of these controversies. 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 in place.

Our analysis focuses on three distinct features of dispute settlement that form the basis of the controversies described above and 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 (SSDS). In the vast majority of investment agreements, by contrast, private investors also have standing (ISDS).

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.

provisions of the USMCA, involving commitments on trade in goods and services and intellectual property rights, rely exclusively on SSDS, as does the WTO.

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 find that for trade agreements, it is never optimal to provide standing to exporters, and consequently that SSDS dominates ESDS. By contrast, for investment agreements we find that standing for foreign investors in the form of ISDS may well be desirable. Factors that bear on the desirability of granting investors standing in investment agreements include the accuracy of the adjudicative process, the political costs faced by the foreign government when initiating a dispute, the probability that a host country will take an action that can be alleged to violate treaty commitments as well as 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. Hence, our analysis provides no support for the inclusion of private standing in trade agreements while it offers qualified – though by no means universal – support for the inclusion of private standing in investment agreements.

Two key observations underlie these results. First, we view trade agreements as concerned primarily with correcting the inefficiencies in market access levels implied by unilateral government policy choices, which means that these are fundamentally *government-to-government* agreements; and we view investment agreements as concerned primarily with solving host government commitment problems with respect to foreign investors, which means that these are fundamentally agreements between the *host government and foreign investors*. Second, we argue that private litigants – exporters and private investors – are imperfect agents for governments and vice versa, and in particular that private litigants would invoke the court more

aggressively than would governments on their behalf.[10]

With these observations in place, our findings on standing can then be understood in two steps. The first step is to observe that as the foreign export industry is an imperfect agent of the foreign government, it does not make sense to give foreign exporters standing in trade disputes when what is at stake is an agreement between the two governments; and as the foreign government is an imperfect agent of foreign investors, it does not make sense to give the foreign government standing in investment disputes when what is at stake is an agreement between the host government and foreign investors. This first step therefore suggests that trade agreements should adopt SSDS while investment agreements should adopt ISDS. The second step is then to observe that foreign complainants never internalize the costs that filing a dispute imposes on the home/host government, and this negative externality leads to a general tendency to overfile relative to efficient litigation levels. Hence, beginning from the position that trade agreements should adopt SSDS, it would never make sense to introduce the possibility of an even more aggressive filer in the form of ESDS; whereas beginning from the position that investment agreements should adopt ISDS, it *might* make sense to restrict standing to a less aggressive filer in the form of SSDS.

The second stage of our analysis concerns the nature of the remedy – the choice between retaliatory sanctions and money damages. Building on our analysis of standing, we now take as given the fact that trade agreements limit standing to governments while investment agreements afford standing to investors. The key trade-off that we exploit in this stage of our analysis is that retaliation is less efficient but that cash damages can be difficult to assess. The optimal remedy in a given setting then depends on which of these forces is stronger in that setting. We demonstrate that it makes sense that investment treaties provide for cash payments while trade agreements do not, provided that the cash value of the profits lost by a foreign investor facing adverse host-country investment policies is sufficiently easy to quantify relative to the cash value of the harm suffered by a foreign government whose exporters face trade protection. And we argue that at a broad level this condition is plausibly met in light of the difficulty of placing a monetary value on the lost jobs and distributional considerations that typically dominate the trade concerns of real-world governments.

[10] This argument assumes that free-rider issues facing exporters and private investors are not too severe. But as we explain below, even severe free-rider issues do not overturn our results on standing, because the relevant thought experiment for standing is whether exporters or private investors should be given standing *in addition* to governments.

The third stage of our analysis adds consideration of the remedial period in the dispute process, that is, the choice between prospective damages only on the one hand, and the addition of retrospective damages on the other. We find that retrospective remedies are 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 we argue is likely to be the case for 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 prospective remedies alone are 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 we argue is likely to be the case for 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 investment treaties. 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 investment treaties (see, for example Konrad, 2017, Horn, 2018, Stahler, 2018, Horn and Tangeras, 2019, Janeba, 2019, Kohler and Stahler, 2019 and Sykes, 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.[11]

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"

[11] 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 (SSDS, but not ESDS), 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 USMCA Chapter 31, 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 "if 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 such treaties 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 investment treaties 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 SSDS or ESDS/ISDS in an optimally designed trade/investment agreement. We first consider the standing issue when the underlying inefficiency to be addressed by the agreement relates to market access/terms of trade issues. Consistent with our discussion in section 2, our formal results are derived in the setting of a trade agreement, reflecting the position that these 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. We then address the standing issue when the underlying inefficiency to be addressed by the

agreement relates to a government's imperfect ability to make commitments to foreign firms that must make sunk investments to serve the domestic market. There our formal results are derived in the setting of an investment agreement, again consistent with our discussion in section 2 and reflecting the dominance of such issues for investment treaties; 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 assumptions in sections 4 and 5 respectively, 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, and where their trade is governed by a trade agreement with 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. We assume that Home gains from protection and that both the Foreign government and its exporting industry suffer from protection in all states of the world. Denoting the Home government's payoff from choosing policy τ in state s as $\omega(\tau; s)$, we then have Home's gain from protection given by $\gamma_G(s) \equiv \omega(P; s) - \omega(FT; s) > 0$ for all s; and similarly, denoting the Foreign agent's payoff as $\omega_f^*(\tau; s)$, where $f \in \{G^*, E^*\}$ depending on whether we consider the government's (G^*) or the exporting industry's (E^*) payoff, we have that Foreign agent f's loss from protection is given by $\gamma_f^*(s) \equiv \omega_f^*(P; s) - \omega_f^*(FT; s) < 0$ 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 in government objectives such as would be reflected by over-weighting producer surplus.

Free trade is the "first best" policy in all states $s \in \sigma^{FT}$ and protection is the "first-best" policy in all states $s \in \sigma^P$, where we define the first-best policy as the policy that maximizes the governments' joint payoff. Letting $\Gamma(s) \equiv \gamma_G(s) + \gamma^*_{G^*}(s)$ denote the governments' joint (positive or negative) gain from protection, we then have 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 Home 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 $\Gamma > 0$.[12] The DSB ruling

[12] 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.

can therefore be thought of as simply a policy determination that maximizes the expected joint payoff of the governments given the DSB signal.[13] 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.[14] 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 constants 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 is realized 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 has no incentive to invoke the DSB and 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 Foreign country to invoke the DSB in response to a policy choice by the Home government of $\tau = P$.

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

[13] See Maggi and Staiger (2011) for a defense of this interpretation of DSB rulings in the context of the GATT/WTO.

[14] 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.

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.[15] 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) 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.*

The content of Lemma 1 is depicted in Figure 2a. Notice that the court has its best impact off-equilibrium, when due to its high accuracy it induces both the Home 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 Home government (in σ^{FT}, if the Home 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 (in σ^{P}, 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 con-

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

tract 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 $\sigma_1^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) < \mu_1^{FT}(s)\right\}$, $\sigma_{2,f}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) \in \left[\mu_1^{FT}(s), \mu_{2,f}^{FT}(s)\right]\right\}$, and $\sigma_{3,f}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) > \mu_{2,f}^{FT}(s)\right\}$, as well as $\sigma_{1,f}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) < \mu_{1,f}^{P}(s)\right\}$, $\sigma_{2,f}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) \in \left[\mu_{1,f}^{P}(s), \mu_2^{P}(s)\right]\right\}$, and, $\sigma_3^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) > \mu_2^{P}(s)\right\}$, we can write:

$$\begin{aligned} L(V_f) \;=\; & \sum_{s \in \sigma_{2,f}^{FT} \cup \sigma_{2,f}^{P}} p(s)\, qk(s)\, |\Gamma(s)| + \\ & \sum_{s \in \sigma_{2,f}^{FT} \cup \sigma_{2,f}^{P}} p(s) \left[c(s) + c_f^*(s)\right] + \\ & \sum_{s \in \sigma_{3,f}^{FT} \cup \sigma_3^{P}} p(s)\, |\Gamma(s)|. \end{aligned} \tag{3.3}$$

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, $s \in \sigma_{2,f}^{FT} \cup \sigma_{2,f}^{P}$. 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, $s \in \sigma_{2,f}^{FT} \cup \sigma_{2,f}^{P}$. 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, $s \in \sigma_{3,f}^{FT} \cup \sigma_3^{P}$.

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.

According to (3.3), the relative merits of SSDS versus ESDS can be evaluated once we specify the complainant's cost of filing $c_f^*(s)$ and payoff from winning in court $\left|\gamma_f^*(s)\right|$ under each choice of standing. 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 $\sigma_{2,f}^{FT}$, $\sigma_{2,f}^{P}$, and $\sigma_{3,f}^{FT}$ through the conditions (3.1) and (3.2).

We assume that the cost of filing is independent of the identity of the complainant. Formally, and recalling that $c_f^*(s) = c_f^* \epsilon^*(s)$, we set the cost of filing for the Foreign exporting industry equal to the cost of filing for the Foreign government with the assumption that

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

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

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

Assumption 2 would hold, for example, in any model where the removal of the Home tariff, by increasing the price of the Foreign export good, would hurt Foreign consumers of that good, or would hurt other Foreign producers who compete for the same factors of production and do not also export the same good to Home. The Foreign government would in some fashion (i.e., with non-zero weight) take all of these effects – both the gains to Foreign exporters *and* the losses to other agents in the Foreign economy – into account when calculating $|\gamma_{G^*}^*(s)|$; but Foreign exporters would ignore the cost imposed on Foreign consumers and other Foreign producers when calculating $|\gamma_{E^*}^*(s)|$. Assumption 2 could also reflect the fact that the Foreign government, unlike the Foreign exporting industry, might apply a "political filter" when evaluating the benefits of winning in court against Home in light of the broader diplomatic relations between the two countries. 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]

[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

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. These sets are illustrated in Figure 2b. The first 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 Foreign 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^*}. The second 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 Home government would choose the inefficient $\tau = P$ with impunity 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 then have:

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

Expression (3.4) highlights the costs and benefits associated with the additional litigation arising under ESDS relative to SSDS. The first line is clearly positive. It captures the fact that the added litigation in states $s \in \Omega_S^P$ is undesirable, since this litigation challenges an efficient policy and thus introduces nothing but court error and litigation costs: $qk(s)\left|\Gamma(s)\right|$ is the expected loss associated with court error, and $c(s) + c^*(s)$ is the joint litigation cost under Assumption 1. The second line is also positive, reflecting the fact that the added litigation in states $s \in \Omega_S^{FT}$ is also undesirable, but the sign is less obvious this time. This is because the added litigation now challenges an *in*efficient policy so that the loss associated with court error and litigation costs is counterbalanced by an efficiency gain $\left|\Gamma(s)\right|$. However, using the filing condition of the Foreign government (equation (3.1) with $f = G^*$), it is easy to show that the loss associated with court error and litigation costs is greater than the efficiency gain.[17] Evidently,

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. In the extreme, this free-rider issue could dilute the filing behavior of the Foreign export industry to the point where it would be a less aggressive filer than the Foreign government, just the opposite of what is implied by 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 exporting industry have standing. And in that more realistic case, the SSDS would handle filings for situations where the free-rider issue caused the Foreign export industry to be a less aggressive filer than the Foreign government, and the ESDS would only be relevant for cases where Assumption 2 applies. Hence, we feel justified in abstracting from this free-rider issue.

[17] In particular, in states $s \in \Omega_S^{FT}$ the Foreign government would not have filed so that $\left[1 - qk(s)\right]\left[-\gamma_{G^*}^*(s)\right] -$

for $s \in \Omega_S^{FT}$ the Foreign *government* does not see a filing as worth the dispute cost while the Home 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.

The two governments would therefore choose to include SSDS rather than ESDS in their trade agreement. Intuitively, this is because under SSDS the Foreign government is itself overly litigious, owing to the negative external effects on the Home government that it does not internalize when making its filing decision; 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 investment treaties , 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, in progress), we exploit the close relationship between exporting and FDI and show that countries have a terms-of-trade motive for

$c^*(s) < 0$. We can then use this inequality to prove that the second line of (3.4) is positive using $[qk(s)|\Gamma(s)| + c(s) + c^*(s)] - |\Gamma(s)| = [1 - qk(s)][\gamma_G(s) + \gamma^*_{G^*}(s)] + c(s) + c^*(s) = [1 - qk(s)]\gamma_G(s) + c(s) - \{[1 - qk(s)][-\gamma^*_{G^*}(s)] - c^*(s)\} > [1 - qk(s)]\gamma_G(s) + c(s) > 0$.

restricting the market access of foreign multinationals just as they do for restricting the market access of foreign exporters: both restrictions reduce the demand for foreign products and thus improve the terms-of-trade. However, this motive is weaker in the context of FDI than in the context of exporting, because local affiliates of foreign multinationals typically also employ local factors which reduces the ability to shift costs to foreign countries. This implies in turn that market access considerations should play less of a role in investment treaties than in trade agreements.[18]

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 consider next the issue of standing in investment agreements, and thus the question whether investment treaties should include SSDS or ISDS. For simplicity and consistent with our discussion in section 2, we assume that investment treaties are only concerned with helping the Home – which we now refer to as Host – government make 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 at 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 can address through the investment treaty. And third, while our reduced-form trade model was sufficient for capturing the market access issues highlighted in the previous subsection, the commitment problem on which we focus here warrants a more explicit modeling of the investment setting

[18] Sykes (2019) draws a similar conclusion.

and the various inefficiencies that can arise. Otherwise, the model essentially follows our earlier structure and we summarize the timing of events in Figure 3.

Finally, as with trade agreements, real-world investment agreements 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." Broadly speaking, we will therefore model the investment treaty as we did the trade agreement, that is, as a vague contract combined with a DSB whose mandate is to interpret the contract when invoked.

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 investor makes a sunk investment I^*, which is converted one-for-one into output through the production function $Q = I^*$. We assume that Host-country 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 $\tilde{P} > 0$.

Host-country 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 is ignored by investors and individual consumers. As in the model of the previous subsection, 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 $e(I^*, s)$. 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. Moreover, 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 turn the ex-post social value of the investment negative – that is, to ensure $PS(I^*) + CS(I^*) - e(I^*, s) < 0$ – for any positive investment level I^*.[19] This assumption simplifies the ensuing analysis by preserving the binary-policy-choice structure of the previous subsection, but it is not necessary for any of the results we emphasize. Under this assumption, the states of the world may be partitioned into those where a negative 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 Foreign 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 Host 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 Host government receives $\bar{\omega}(I^*, T, s) = 0$ if there is a taking and the output from the investment is destroyed, and if there is not a taking the Host government suffers the negative payoff $\bar{\omega}(I^*, FT, s) = CS(I^*) - e(I^*, s) < 0$.[20] For $s \in \sigma^{FT}$, the Host government receives $\bar{\omega}(I^*, T, s) = CS(I^*) + \kappa PS(I^*)$ if there is a taking and receives $\bar{\omega}(I^*, FT, s) = CS(I^*)$ 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}.[21] In particular, if κ were equal to one then the

[19] 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.

[20] Here and throughout, we use "over-bars" to distinguish notationally between functions in our analysis of investment agreements that also appear in our analysis of trade agreements.

[21] 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 inefficiency facing the Host government and to the potential role of an investment treaty in helping to address that inefficiency.

We can define the ex-post gain that the Host government enjoys from a taking in state s as

$$\bar{\gamma}_G(I^*, s) \equiv \bar{\omega}(I^*, T, s) - \bar{\omega}(I^*, FT, s) = \begin{cases} e(I^*, s) - CS(I^*) & \text{for } s \in \sigma^T \\ \kappa PS(I^*) & \text{for } s \in \sigma^{FT} \end{cases} \tag{3.5}$$

and the lost rents suffered by the Foreign investor in a taking as

$$\bar{\gamma}^*_{I^*}(I^*) \equiv -PS(I^*) \quad \text{for all } s. \tag{3.6}$$

The joint ex-post gain from a taking for the Host government and Foreign investor is then given by $\bar{\Gamma}(I^*, s) \equiv \bar{\gamma}_G(I^*, s) + \bar{\gamma}^*_{I^*}(I^*)$, and we have

$$\bar{\Gamma}(I^*, s) = \begin{cases} -[PS(I^*) + CS(I^*) - e(I^*, s)] > 0 & \text{for } s \in \sigma^T \\ -(1-\kappa) PS(I^*) < 0 & \text{for } s \in \sigma^{FT}. \end{cases} \tag{3.7}$$

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 when I^* is taken as given, we have $\iota_{FB} = FT$ for $s \in \sigma^{FT}$ and $\iota_{FB} = T$ for $s \in \sigma^T$.

We emphasize that our 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.

ex-post efficient.

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 policy choice would correspond to 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 in the Host country, 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 these 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 $s \in \sigma^{FT}$ is realized (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

$$p^{FT}\bar{p}\tilde{P}\left(I_C^*\right) = r^*, \tag{3.8}$$

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

$$\begin{aligned} E_s\left[\bar{\omega}\left(I_C^*, \iota_C, s\right)\right] &= p^{FT}\left[CS\left(I_C^*\right) + \left(1-\bar{p}\right)\kappa PS\left(I_C^*\right)\right] \\ &= p^{FT}\left[CS\left(I_C^*\right) + PS\left(I_C^*\right) - \left(1-\bar{p}\right)\left(1-\kappa\right)PS\left(I_C^*\right)\right] - r^* I_C^*, \end{aligned} \tag{3.9}$$

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 $\frac{\partial E_s\left[\bar{\omega}\left(I_C^*,\iota_C,s\right)\right]}{\partial \bar{p}} > 0$ for $\bar{p} \in [0,1)$ by differentiating (3.8) and (3.9), 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. For future reference, we define the first-best investment level, I_{FB}^*, as the investment level solving (3.8) when $\bar{p}=1$.

The commitment problem that arises with $\bar{p}<1$ implies 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 an investment treaty. 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 investment treaty 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 and in contrast to our analysis of trade agreements in the previous subsection, in the absence of an investment treaty there is no inefficiency that can be traced to a *government-to-government international policy externality*: instead, the commitment problem at the heart of investment agreements is a *government-to-investor policy commitment problem* that creates a Host-country (domestic) inefficiency, the costs of which are borne entirely by the Host government. What an investment treaty can do is serve as a way for the Host government to make policy commitments to Foreign investors ex ante, and thereby address this domestic inefficiency. And this has an important implication: we can evaluate the investment treaty based on how close it comes to achieving $E_s\left[\bar{\omega}\left(I_{FB}^*,\iota_{FB},s\right)\right]$, 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

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

With probability p^{FT}, the Foreign production 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 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 to induce under an up-front investment incentive program, $\hat{I}^*$, the Host government solves $\max_{I^*} E_s\left[\omega\left(I^*,\iota_C,s\right)\right]$. This yields the first-order condition

$$\tilde{P}\left(\hat{I}^*\right)=\frac{r^*}{p^{FT}}+\left(1-\bar{p}\right)\left(1-\kappa\right)\frac{\partial PS\left(\hat{I}^*\right)}{\partial I^*}, \tag{3.10}$$

which implicitly defines $\hat{I}^*$. It is easy to verify using equations (3.8) and (3.10) that $\hat{I}^*$ 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

$$\begin{aligned} E_s\left[\bar{\omega}\left(\hat{I}^*,\iota_C,s\right)\right] &= p^{FT}\left[CS\left(\hat{I}^*\right)+PS\left(\hat{I}^*\right)-\left(1-\bar{p}\right)\left(1-\kappa\right)PS\left(\hat{I}^*\right)\right]-r^*\hat{I}^* \\ &= p^{FT}\left[CS\left(\hat{I}^*\right)+\left(1-\bar{p}\right)\left(1-\kappa\right)\frac{\partial\tilde{P}\left(\hat{I}^*\right)}{\partial I^*}\left(\hat{I}^*\right)^2\right]. \end{aligned} \tag{3.11}$$

Equations (3.8)-(3.11) 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 Host government then optimally chooses $\hat{I}^*>I^*_C$. It is also easy to see from these equations that this same condition

ensures that the up-front investment program cannot achieve the first-best outcome since:

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

We summarize with:

Lemma 2. *As long as* $(1-\bar{p})(1-\kappa) > 0$*, an up-front investment incentive program can help solve the Host-government commitment problem with regard to Foreign investors, but it cannot 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.[22] Intuitively, this is because such programs can address the ex-ante inefficiency due to underinvestment, but they cannot address the ex-post inefficiency resulting from takings in states $s \in \sigma^{FT}$. The possibility of correcting these latter inefficiencies provides a natural role for an investment treaty. We next turn to assessing this role, and evaluating whether it is best served by an investment treaty that features ISDS or rather SSDS.

Investment Agreements We are now ready to dive into our full analysis allowing for both up-front investment incentives and an investment agreement. The investment agreement determines what happens once the investment is sunk and works exactly like our trade agreement from before. In particular, the Host government chooses its investment policy; if it chooses a taking, the Foreign investor (in the case of ISDS) or Foreign government (in the case of SSDS) then decides whether to file a complaint with the DSB; and if a complaint is filed, the DSB then issues a ruling based on a noisy signal of which policy is first-best in the realized state of the world. The up-front investment incentive determines the ex-ante investment, and the Host government chooses the level of the investment incentive to maximize its ex-ante expected payoff

[22] 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). Nevertheless, 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 inefficiency.

keeping the implications of the investment agreement in mind. We proceed by backwards induction, considering first the Foreign agent's filing choice, then the Host government's investment policy choice, and finally the Host government's up-front investment incentive choice.

Recall that for the Host government, the ex-post gain from a taking, $\bar{\gamma}_G(I^*, s)$, is defined in (3.5) and note that it is positive in all states of the world, while the loss to the Foreign investor is defined in (3.6) and given by $\bar{\gamma}^*_{I^*}(I^*) = -PS(I^*)$. Similarly to our earlier analysis of trade agreements, here we allow the Foreign government's loss from a taking (and hence payoff from winning in court) to differ from that of the Foreign investor, and we capture the loss from a taking for each of these potential Foreign claimants with the simple parameterization $\bar{\gamma}^*_f(I^*) \equiv -\bar{\gamma}^*_f PS(I^*)$ for $f \in \{G^*, I^*\}$, where the parameter $\bar{\gamma}^*_{I^*} \equiv 1$ by (3.6) and where we will later place restrictions on the parameter $\bar{\gamma}^*_{G^*}$ but for now only assume that it is positive. We denote by $\bar{c}^*_f(\cdot)$ the cost incurred by Foreign complainant $f \in \{G^*, I^*\}$ whenever it invokes the DSB, and we allow this cost to be a function of the level of 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 $\bar{c}^*_f(I^*, s) \equiv \left[\bar{c}^*_f \epsilon^*(s)\right] PS(I^*)$. We make the analogous assumption for the Host government: if the Host government is taken to court in state s, it incurs a cost $\bar{c}(I^*, s) \equiv [\bar{c}\epsilon(s)]\,\bar{\gamma}_G(I^*, s)$ to defend the taking.[23] As with our earlier analysis of trade agreements, we will later introduce assumptions about how the cost of filing and payoff from winning in court varies across complainants and explicitly consider equilibrium outcomes under ISDS and SSDS, but for now we develop the model for general complainant $f \in \{G^*, I^*\}$.

As before, we assume that the realized state s is observed by all agents including the DSB, and that Γ is observed by the agents but not by the DSB.[24] 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 ex-post (once-the-foreign-investment-is-sunk) joint payoff of the Host government and the Foreign investor given its noisy signal of Γ.

[23] As will become clear below, allowing dispute costs to rise with the level of foreign investment in this way ensures that the Foreign filing decision and Host policy choice are independent of the level of investment I^*, simplifying the analysis to follow.

[24] 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 facility is shut down.

It is worth pausing here to elaborate on what we have in mind more broadly regarding the DSB ruling. Under the interpretation that T represents a "regulatory taking," the DSB ruling could be seen in a richer model 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 outright 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 investor 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.[25]

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

$$\Pr(\text{DSB ruling is } FT \mid s) \times |\bar{\gamma}_f^*(I^*)| > \bar{c}_f^*(I^*, s). \tag{3.12}$$

Condition (3.12) 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 (3.12)

[25] More specifically, in the case of outright 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 see how our model maps over to this case, and in analogy with our simplification of two policies T and FT, suppose 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 suppose 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).

fails – because then $\iota = T$ can be set without triggering a dispute – or if (3.12) holds and the expected benefit to the Host government from a taking exceeds the cost to the Host government of a dispute:

$$\Pr(\text{DSB ruling is } T \mid s) \times \bar{\gamma}_G(I^*, s) > \bar{c}(I^*, s). \tag{3.13}$$

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 $\bar{\mu}_1^{FT}(s) \equiv \bar{c}\epsilon(s)$, $\bar{\mu}_{2,f}^{FT}(s) \equiv 1 - \frac{\bar{c}_f^*}{\bar{\gamma}_f^*}\epsilon^*(s)$, $\bar{\mu}_{1,f}^{T}(s) \equiv \frac{\bar{c}_f^*}{\bar{\gamma}_f^*}\epsilon^*(s)$, and $\bar{\mu}_2^T(s) \equiv 1 - \bar{c}\epsilon(s)$, and noting 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 (3.12) and (3.13) immediately imply the following result:

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.*

Note that Lemma 3 is analogous to Lemma 1 if the Host government is unconstrained by domestic institutions, in the sense that the equilibrium actions follow an intuitive sorting

along the dimension of DSB quality. If the DSB quality is high, the Host government makes the efficient policy choice and there is no dispute; if the DSB quality is intermediate, the Host government chooses $\iota = T$ and there is a dispute; and if the DSB quality is low, the Host government chooses the inefficient policy and there is no dispute. The only substantive difference between Lemma 3 and Lemma 1 is that the Host government is now constrained by domestic institutions with probability $\bar{p}$ which forces it to implement the efficient policy no matter what. For future reference, we again define 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, $\bar{\sigma}_{3}^{T} \equiv \left\{s \in \bar{\sigma}^{T} \mid qk(s) > \bar{\mu}_{2}^{T}(s)\right\}$.

With Lemma 3 in hand, we can now roll back to the first stage and solve for the optimal investment incentives offered by the Host government. The Host government can induce any level of investment I^* by offering up-front investment incentives of $\left\{r^* - E_s\left[\rho_f^*(I^*, s)\right]\right\} I^*$, where $E_s\left[\rho_f^*(I^*, s)\right]$ is the expected return to the Foreign investor on an investment of I^* under an investment treaty with standing choice $f \in \{G^*, E^*\}$. In the Appendix, we provide the expression for $E_s\left[\rho_f^*(I^*, s)\right]$ and use it to show that the expected welfare of the Host government associated with Foreign investment choice I^* is given by:

$$
\begin{aligned}
E_s\left[\bar{\omega}_f(I^*, s)\right] \;=\; & (1-\bar{p}) \sum_{s \in \bar{\sigma}_1^{FT}} p(s) \left[CS(I^*) + PS(I^*)\right] \\
& + (1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,f}^{FT}} p(s) \left\{CS(I^*) + PS(I^*) - qk(s)(1-\kappa) PS(I^*) - \bar{c}(I^*, s) - \bar{c}_f^*(I^*, s)\right\} \\
& + (1-\bar{p}) \sum_{s \in \bar{\sigma}_{3,f}^{FT}} p(s) \left[CS(I^*) + \kappa PS(I^*)\right] \\
& - (1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,f}^{T}} p(s) \left\{qk(s)\left[e(I^*, s) - CS(I^*) - PS(I^*)\right] + \bar{c}(I^*, s) + \bar{c}_f^*(I^*, s)\right\} \\
& - (1-\bar{p}) \sum_{s \in \bar{\sigma}_{3}^{T}} p(s) \left[e(I^*, s) - CS(I^*) - PS(I^*)\right] \\
& + \bar{p} p^{FT} \left[CS(I^*) + PS(I^*)\right] \\
& - \bar{p} \sum_{s \in \bar{\sigma}_{2,f}^{T} \cup \bar{\sigma}_{3}^{T}} p(s) \left\{qk(s)\left[e(I^*, s) - CS(I^*) - PS(I^*)\right] + \bar{c}(I^*, s) + \bar{c}_f^*(I^*, s)\right\} \\
& - r^* I^*.
\end{aligned}
\tag{3.14}
$$

The interpretation of this expression is intuitive, once it is understood that the Host gov-

ernment must pay the Foreign investor the amount r^*I^* in equilibrium, as reflected in the last line of (3.14); 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 an investment treaty with standing choice $f \in \{I^*, G^*\}$ as characterized above. The first five lines of (3.14) 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 policy-making (which happens with probability $1-\bar{p}$). And the next two lines add the surplus that is generated if the Host government is constrained to implement the first-best policies (which happens with probability $\bar{p}$). To see this, note that the efficient policy yields a surplus for the Host government of $CS(I^*) + PS(I^*)$ in states $s \in \sigma^{FT}$ and 0 in states $s \in \sigma^T$, the inefficient policy yields a surplus for the Host government of $CS(I^*) + \kappa PS(I^*)$ in states $s \in \sigma^{FT}$ and $CS(I^*) + PS(I^*) - e(I^*, s) < 0$ in states $s \in \sigma^T$, the DSB makes a mistake with probability $qk(s)$, and invoking the DSB costs $c_f^*(I^*, s) + c(I^*, s)$.

Using the expression for $E_s[\bar{\omega}_f(I^*, s)]$ in (3.14), we can now solve for the optimal level of FDI in the presence of an investment treaty with standing choice $f \in \{I^*, G^*\}$, which we denote by $\bar{I}_f^*$, defined implicitly by $\frac{\partial E_s[\bar{\omega}_f(\bar{I}_f^*, s)]}{\partial I^*} = 0$. The associated first order condition can be manipulated to yield the following implicit characterization of $\bar{I}_f^*$:

$$
\begin{aligned}
\tilde{P}\left(\bar{I}_f^*\right) \;=\; & \frac{r^*}{p^{FT}} \\
& + (1-\bar{p})(1-\kappa)\frac{\sum_{s\in\bar{\sigma}_{2,f}^{FT}} p(s)\,qk(s) + p_{3,f}^{FT}}{p^{FT}}\frac{\partial PS\left(\bar{I}_f^*\right)}{\partial I^*} \\
& + (1-\bar{p})\sum_{s\in\bar{\sigma}_{2,f}^{T},\bar{\sigma}_2^{FT}}\frac{p(s)}{p^{FT}}\left[\frac{\partial \bar{c}\left(\bar{I}_f^*, s\right)}{\partial I^*} + \frac{\partial \bar{c}_f^*\left(\bar{I}_f^*, s\right)}{\partial I^*}\right] \\
& + (1-\bar{p})\left[\sum_{s\in\bar{\sigma}_{2,f}^{T}}\frac{p(s)\,qk(s)}{p^{FT}} + \sum_{s\in\bar{\sigma}_3^T}\frac{p(s)}{p^{FT}}\right]\left[e(s) - \tilde{P}\left(\bar{I}_f^*\right)\right] \\
& + \bar{p}\sum_{s\in\bar{\sigma}_{2,f}^{T},\bar{\sigma}_3^T}\frac{p(s)}{p^{FT}}\left[\frac{\partial \bar{c}\left(\bar{I}_f^*, s\right)}{\partial I^*} + \frac{\partial \bar{c}_f^*\left(\bar{I}_f^*, s\right)}{\partial I^*}\right] \\
& + \bar{p}\sum_{s\in\bar{\sigma}_{2,f}^{T},\bar{\sigma}_3^T}\frac{p(s)\,qk(s)}{p^{FT}}\left[e(s) - \tilde{P}\left(\bar{I}_f^*\right)\right].
\end{aligned}
\tag{3.15}
$$

We can see from the terms on the right-hand-side of (3.15) that there are a number of forces that determine the optimal level of investment $\bar{I}_f^*$ in the presence of an investment treaty 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 (3.15), then this expression would imply $\bar{I}_f^* = I_{FB}^*$. The terms on the remaining lines then capture various inefficiencies that make $\bar{I}_f^* < I_{FB}^*$, 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 (3.10) but down-weighted by the reduced probability of a taking in σ^{FT} under the investment treaty. 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 fourth line reflects the expected welfare costs that arise when efficient takings are prevented, either by an incorrect DSB ruling or in the shadow of a low-quality DSB. 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 agent f 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.[26]

Finally, plugging $\bar{I}_f^*$ into the expression for $E_s\left[\bar{\omega}_f\left(I^*, s\right)\right]$ in (3.14) and simplifying yields

$$\begin{aligned} E_s\left[\bar{\omega}_f\left(\bar{I}_f^*, s\right)\right] &= (1-\bar{p})\left[p^{FT} + p_3^T + \sum_{s\in\bar{\sigma}_{2,f}^T} p\left(s\right) qk\left(s\right)\right] CS\left(\bar{I}_f^*\right) \\ &\quad + (1-\bar{p})(1-\kappa)\left[\sum_{s\in\bar{\sigma}_{2,f}^{FT}} p\left(s\right) qk\left(s\right) + p_{3,f}^{FT}\right]\left(\bar{I}_f^*\right)^2 \frac{\partial \tilde{P}\left(\bar{I}_f^*\right)}{\partial \bar{I}_f^*} \\ &\quad - (1-\bar{p}) \sum_{s\in\bar{\sigma}_{2,f}^{FT}\cup\bar{\sigma}_{2,f}^T} p\left(s\right)\left\{\bar{c}\left(\bar{I}_f^*, s\right) + \bar{c}_f^*\left(\bar{I}_f^*, s\right) - \bar{I}_f^*\left[\frac{\partial \bar{c}\left(\bar{I}_f^*, s\right)}{\partial I^*} + \frac{\partial \bar{c}_f^*\left(\bar{I}_f^*, s\right)}{\partial I^*}\right]\right\} \\ &\quad + \bar{p}\left[p^{FT} + \sum_{s\in\bar{\sigma}_{2,f}^T\cup\bar{\sigma}_3^T} p\left(s\right) qk\left(s\right)\right] CS\left(I^*\right) \\ &\quad - \bar{p} \sum_{s\in\bar{\sigma}_{2,f}^T\cup\bar{\sigma}_3^T} p\left(s\right)\left\{\bar{c}\left(\bar{I}_f^*, s\right) + \bar{c}_f^*\left(\bar{I}_f^*, s\right) - \bar{I}_f^*\left[\frac{\partial \bar{c}\left(\bar{I}_f^*, s\right)}{\partial I^*} + \frac{\partial \bar{c}_f^*\left(\bar{I}_f^*, s\right)}{\partial I^*}\right]\right\}. \end{aligned} \tag{3.16}$$

This expression reveals that the investment treaty only improves upon the program of up-front investment incentives if court quality is sufficiently high and domestic institutions are suffi-

[26] We have assumed for simplicity that the probability of court mistakes $qk(s)$ is independent of $\bar{p}$, the probability that the Host government has its hands tied to implement the first-best policy, but we could allow the probability of court mistakes to depend on $\bar{p}$ without altering in a substantial way any of our results.

ciently weak. If $qk(s) > \max\left\{\bar{\mu}_{2,f}^{FT}(s), \bar{\mu}_2^T(s)\right\}$ for all s so that the sets $\left\{\bar{\sigma}_1^{FT}, \bar{\sigma}_{2,f}^{FT}, \bar{\sigma}_{1,f}^T, \bar{\sigma}_{2,f}^T\right\}$ are empty, the court quality is then so low that the inefficient policy is always implemented without any dispute. In this case we have using (3.11) and (3.14) that $E_s\left[\bar{\omega}_f\left(\bar{I}_f^*, s\right)\right] \leq E_s\left[\bar{\omega}\left(\hat{I}^*, \iota_C, s\right)\right]$ so that the investment treaty makes things worse. Also, if $\bar{p} = 1$ so that domestic institutions alone would solve the Host government commitment problem and $E_s\left[\bar{\omega}\left(\hat{I}^*, \iota_C, s\right)\right] = E_s\left[\bar{\omega}\left(I_{FB}^*, \iota_{FB}, s\right)\right]$, then the investment treaty cannot make things better and makes things worse as long as $qk(s) > \bar{\mu}_1^T(s)$ for some s and there is some litigation in equilibrium. On the other hand, if $qk(s) < \min\left\{\bar{\mu}_1^{FT}(s), \bar{\mu}_1^T(s)\right\}$ for all s so that 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, the court quality is so high that the efficient policy is always implemented without any dispute and thus $E_s\left[\bar{\omega}_f\left(\bar{I}_f^*, s\right)\right] = E_s\left[\bar{\omega}\left(I_{FB}^*, \iota_{FB}, s\right)\right] > E_s\left[\bar{\omega}\left(\hat{I}^*, \iota_C, s\right)\right]$.[27]

We summarize with:

Proposition 2. *The introduction of an investment treaty 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.*

The possibility described by Proposition 2, that a purely unilateral intervention could dominate an international agreement, does not arise for the case of trade agreements. This difference reflects the different nature of the inefficiencies addressed across the two kinds of agreements, a fundamentally domestic inefficiency in the case of investment treaties 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 an investment treaty to improve upon up-front investment incentives by themselves, at least when standing is optimally allocated in the investment treaty.

SSDS versus ISDS We now evaluate the desirability of adopting SSDS versus ISDS. This again requires us to take a stance on the complainant's cost of filing and payoff from winning in court under each choice of standing, and we impose analogous assumptions to our earlier

[27] We are ignoring the issues associated with the endogeneity of contract choice and court mandate to the quality of the court that are highlighted in the analysis of Maggi and Staiger (2011). But the point we emphasize here 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.

Assumption 1 and Assumption 2. In particular, we assume that the cost of filing for the Foreign government is the same as the cost of filing for the Foreign investor, namely

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

implying $\bar{c}^*_{G^*}(I^*, s) = \bar{c}^*_{I^*}(I^*, s) = [\bar{c}^* \epsilon^*(s)] PS(I^*) \equiv \bar{c}^*(I^*, s)$ for all s. And we assume that the Foreign investor suffers *more* from a taking than the Foreign government in the sense that

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

which amounts to the parameter restriction $\bar{\gamma}^*_{G^*} < \bar{\gamma}^*_{I^*}$ since then $|\bar{\gamma}^*_{G^*}(I^*)| = \bar{\gamma}^*_{G^*} PS(I^*) < PS(I^*) = |\bar{\gamma}^*_{I^*}(I^*)|$.

Since the Host country is assumed to be small so that its choice of investment policy does not affect the world interest rate r^* or any other international prices, we cannot justify Assumption 2′ on the same economic grounds as our earlier Assumption 2. But we can again invoke the argument that the Foreign government applies a "political filter," which takes into account the broader political, diplomatic and public relations repercussions of winning in court against Home.[28] In the end, Assumption 1′ and Assumption 2′ again imply that adopting ISDS rather than SSDS amounts to the Foreign government delegating filing decisions to a more aggressive filer than itself.

Given Assumption 1′ and Assumption 2′, are there conditions (model parameter ranges) under which ISDS could be part of an optimally designed investment treaty? 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 policy regime complicate such comparisons in the context of investment treaties. Instead we

[28] 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.

hold investment at the optimal level under an investment treaty with SSDS ($\bar{I}^*_{G^*}$) as characterized by (3.15), and we use (3.14) to calculate $E_s[\bar{\omega}_{I^*}(\bar{I}^*_{G^*}, s)] - E_s[\bar{\omega}_{G^*}(\bar{I}^*_{G^*}, s)]$, seeking conditions under which this difference is positive, which then provide sufficient conditions for $E_s[\bar{\omega}_{I^*}(\bar{I}^*_{I^*}, s)] > E_s[\bar{\omega}_{G^*}(\bar{I}^*_{G^*}, s)]$ and hence for ISDS to be part of an optimal investment treaty, given that $E_s[\bar{\omega}_{I^*}(\bar{I}^*_{I^*}, s)] \geq E_s[\bar{\omega}_{I^*}(\bar{I}^*_{G^*}, s)]$.

To express $E_s[\bar{\omega}_{I^*}(\bar{I}^*_{G^*}, s)] - E_s[\bar{\omega}_{G^*}(\bar{I}^*_{G^*}, s)]$, we proceed as above and first define the sets $\bar{\Omega}^{FT} \equiv \left\{ s \in \sigma^{FT} \mid qk\left(s\right) \in \left[\bar{\mu}^{FT}_{2,G^*}\left(s\right), \bar{\mu}^{FT}_{2,I^*}\left(s\right)\right] \right\}$ and $\bar{\Omega}^{T} \equiv \left\{ s \in \sigma^{T} \mid qk\left(s\right) \in \left[\bar{\mu}^{T}_{1,I^*}\left(s\right), \bar{\mu}^{T}_{1,G^*}\left(s\right)\right] \right\}$. 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 (3.14), we can now write $\Delta_{BIT} \equiv E_s[\bar{\omega}_{I^*}(\bar{I}^*_{G^*}, s)] - E_s[\bar{\omega}_{G^*}(\bar{I}^*_{G^*}, s)]$ as

$$\begin{aligned}
\Delta_{BIT} \;=\; & -\sum_{s\in\bar{\Omega}^{T}} p\left(s\right)\left\{qk\left(s\right)\left[e\left(\bar{I}^*_{G^*}, s\right) - CS\left(\bar{I}^*_{G^*}\right) - PS\left(\bar{I}^*_{G^*}\right)\right]\right\} \\
& -\sum_{s\in\bar{\Omega}^{T}} p\left(s\right)\left\{\bar{c}\left(\bar{I}^*_{G^*}, s\right) + \bar{c}^*\left(\bar{I}^*_{G^*}, s\right)\right\} \\
& +\left(1-\bar{p}\right)\sum_{s\in\bar{\Omega}^{FT}} p\left(s\right)\left\{\left(1-qk\left(s\right)\right)\left(1-\kappa\right)PS\left(\bar{I}^*_{G^*}\right) - \bar{c}\left(\bar{I}^*_{G^*}, s\right) - \bar{c}^*\left(\bar{I}^*_{G^*}, s\right)\right\}.
\end{aligned} \tag{3.17}$$

Expression (3.17) summarizes the costs and benefits associated with the additional litigation arising under ISDS relative to SSDS. It is analogous to our earlier expression (3.4) pertaining to trade agreements, except that (3.17) expresses net gains rather than net losses (and thus has the opposite sign). The first line of (3.17) is negative since $e\left(\bar{I}^*_{G^*}, s\right) - CS\left(\bar{I}^*_{G^*}\right) - PS\left(\bar{I}^*_{G^*}\right) > 0$ in states $s \in \bar{\Omega}^{T}$, as is the second line. These two lines argue against ISDS, reflecting the fact that the extra litigation induced under ISDS relative to SSDS is undesirable in states $s \in \bar{\Omega}^{T}$, since this litigation challenges an efficient policy and thus introduces nothing but court error and litigation costs. But the sign of the third line in (3.17) is ambiguous. This can be confirmed by rearranging terms and rewriting this line in the equivalent form

$$\left(1-\bar{p}\right)\sum_{s\in\bar{\Omega}^{FT}} p\left(s\right)\left\{\left[\left(1-qk\left(s\right)\right)PS\left(\bar{I}^*_{G^*}\right) - \bar{c}^*\left(\bar{I}^*_{G^*}, s\right)\right] - \left[\left(1-qk\left(s\right)\right)\kappa PS\left(\bar{I}^*_{G^*}\right) + \bar{c}\left(\bar{I}^*_{G^*}, s\right)\right]\right\},$$

and noting that the first term in square brackets must be positive by the filing condition for Foreign investors in the set of states $s \in \bar{\Omega}^{FT}$, which recall are states in σ^{FT} where the Host

government would implement a taking with impunity under SSDS but would face litigation under ISDS. Evidently, this expression will be positive if κ and c are sufficiently small so that the second term in square brackets is sufficiently small. And the set of states over which this expression is summed, $\bar{\Omega}^{FT}$, will be larger the smaller is $\bar{\gamma}^*_{G^*}$, i.e., the greater the divergence is between the Foreign government and Foreign investors in the payoff to filing according to Assumption 2′. Finally, note that the first two lines of (3.17) can be set arbitrarily close to zero by reducing the probability that states $s \in \sigma^T$ occur, $p^T \equiv \sum_{s \in \sigma^T} p(s)$. Hence, with the third line of (3.17) guaranteed to be positive if κ, $\bar{c}$ and $\bar{\gamma}^*_{G^*}$ are sufficiently low, ISDS dominates SSDS if p^T is also sufficiently low so that it is sufficiently rare for a taking to be socially efficient.

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 an investment treaty. In reality, the relevant design choice is more aptly described as whether to include ISDS in an investment treaty 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 investment treaty depends on a number of subtle tradeoffs. But if the foreign government faces high political costs of initiating a dispute, 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 expropriation is socially efficient only in unusual circumstances, then it is optimal to give investors standing to bring disputes in an investment treaty. That is, if $\bar{\gamma}^*_{G^*}$, κ, $\bar{c}$, and p^T are sufficiently low, an optimally designed investment treaty should include both an SSDS and an ISDS.*

Summarizing, we can conclude from Propositions 1 and 3 that, in contrast to trade agreements, in the context of investment agreements there is a case to be made for going beyond SSDS and including standing for private agents in the form of ISDS, though even in this context the case for ISDS is far from absolute. At a broad level, the intuition for these conclusions can be understood in two steps. A first step follows from our earlier observation that a trade agreement is a contract between *governments* to address a government-to-government international policy externality, while an investment agreement is a contract between a Host government and *foreign investors* to address a government-to-investor policy commitment problem. So, if the

Foreign export industry is an imperfect agent of the Foreign government (which we impose in Assumption 2), it does not make sense to give Foreign exporters standing in trade disputes; and if the Foreign government is an imperfect agent of the Foreign investor (which we impose in Assumption 2′), it does not make sense to give the Foreign government standing in investment disputes. This first step therefore suggests that trade agreements should adopt SSDS while investment agreements should adopt ISDS. The second step of the intuition is then to observe that Foreign complainants never internalize the costs that filing a dispute imposes on the Home/Host government, and this negative externality leads to a general tendency to overfile relative to efficient litigation levels. Hence, beginning from the position that trade agreements should adopt SSDS, it would never make sense to introduce the possibility of an even more aggressive filer in the form of ESDS (as is implied by Assumption 2); whereas beginning from the position that investment agreements should adopt ISDS, it *might* make sense to restrict standing to a less aggressive filer in the form of SSDS (as implied by Assumption 2′).

Standing for disputes over commitments to investors more generally We have analyzed commitment issues with respect to foreign investors within the context of investment treaties. 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.[29] In principle, our analysis above (with the interpretation now excluding the possibility of outright 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 investment treaties.

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 investment treaties, because 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 outright 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

[29] 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.

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 investment treaties than it is in the context of trade agreements. And if setting up ISDS involves some fixed-cost component so that it is not worth doing below some minimal level of hold-up threat, then this distinction could account for the inclusion of ISDS provisions in investment treaties when the conditions of Proposition 3 are satisfied but no analogous inclusion of ESDS provisions in trade agreements to handle commitment issues there. 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 even when those agreements are designed to address important commitment issues.

4. Nature of the Remedy

We have assumed thus far that when a case is filed and the court sides with the complainant, the defendant has no choice but to "cease and desist" the policy that has been found to be illegal. In reality, however, convicted violators of trade and investment agreements 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 across trade and investment agreements which is the source of much public controversy: while trade agreements allow the injured party to engage in reciprocal retaliation, investment treaties provide explicitly for cash payments.

In this section, we extend our baseline models of trade and investment agreements from section 3 to explore this difference in the nature of the remedy. 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 choose instead to maintain $\tau = P$ or $\iota = T$ and make damage payments. We compare two forms of damage payment: one in which the court allows the injured party to engage in reciprocal retaliation, and another in which the injured party is awarded cash damages by the court. The key trade-off featured by our extended models is that retaliation is less efficient but that cash damages can be difficult to assess. Which remedy is optimal then depends on which force is stronger. We demonstrate that it makes sense that investment treaties 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. And we argue that at a broad level this condition is plausibly met in practice.

To make our points as clearly as possible, we build on our analysis of standing and now take as given that trade agreements limit standing to governments while investment agreements afford standing to investors. And we adopt two further simplifying assumptions in this section. First, when retaliation is the remedy we assume that retaliation is sufficiently inefficient and costly that the Home or Host government switches to $\tau = FT$ or $\iota = FT$ when convicted by the DSB to avoid retaliation - this means that the analysis of trade and investment agreements with retaliation is exactly the same as in our earlier baseline analysis of these agreements in section 3 (with the trade agreement featuring SSDS and the investment treaty featuring ISDS). Second, when cash payments are the remedy, we assume that cash payments are perfectly efficient so that surplus can be costlessly transferred internationally. With these two simplifying assumptions we adopt an extreme position on the inefficiency of retaliation relative to cash, so that we can focus our analysis of the optimal remedy in trade and investment agreements on the degree of difficulty faced by the court in assessing damages in each setting.

4.1. Trade Agreements

We first consider the choice of retaliation versus cash payments in the context of trade agreements. As indicated above, we assume that the foreign government has standing in trade disputes, and we take retaliation to be a sufficiently inefficient form of damage payment that under retaliation our earlier analysis of trade agreements with SSDS still applies. Our task is then to introduce cash damages into our model of trade agreements, and compare the outcomes from our earlier model (which we now refer to as the outcomes under the V_R *institution*) to the outcomes under cash damages derived here (which we refer to as the outcomes under the V_C *institution*).

To capture the notion that the court may struggle 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 $d^*(s)$, we assume $\Pr[d^*(s) = |\gamma^*_{G^*}(s)|] = 1 - 2m(s)$, $\Pr[d^*(s) > |\gamma^*_{G^*}(s)|] = m(s)$, and $\Pr[d^*(s) < |\gamma^*_{G^*}(s)|] = m(s)$, with $m(s) \in [0, \frac{1}{2}]$. Hence, the court awards the correct damages with probability $1 - 2m(s)$ and overes-

timates or underestimates the damages with symmetric probabilities $m(s)$. Defining $d^*_{high} \equiv E[d^*(s) \mid d^*(s) > |\gamma^*_{G^*}(s)|]$ and $d^*_{low} \equiv E[d^*(s) \mid d^*(s) < |\gamma^*_{G^*}(s)|]$, we further assume that $d^*_{high} > \gamma_G(s)$ and $d^*_{low} < \gamma_G(s)$ for all s. So, if the court overestimates the damages, Home can expect to be assessed damages in the amount of d^*_{high}, 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 d^*_{low} 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 ($d^*(s)$), 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

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

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

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

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 $\mu^{FT}_{1,C}(s) \equiv \frac{c(s)-m(s)[\gamma_G(s)-d^*_{low}]}{\gamma_G(s)-m(s)[\gamma_G(s)-d^*_{low}]}$, $\mu^{FT}_{2,C}(s) \equiv 1 - \frac{c^*(s)}{|\gamma^*_{G^*}(s)|-m(s)[|\gamma^*_{G^*}(s)|-d^*_{low}]}$, $\mu^{P}_{1,C}(s) \equiv \frac{c^*(s)}{|\gamma^*_{G^*}(s)|-m(s)[|\gamma^*_{G^*}(s)|-d^*_{low}]}$, and $\mu^{P}_{2,C}(s) \equiv \frac{\gamma_G(s)-c(s)}{\gamma_G(s)-[1-2m(s)][\gamma_G(s)-|\gamma^*_{G^*}(s)|]-m(s)[\gamma_G(s)-d^*_{low}]}$,

and noting as before that $\mu_{1,C}^{FT}(s) < \mu_{2,C}^{FT}(s)$ and $\mu_{1,C}^{P}(s) < \mu_{2,C}^{P}(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (4.1) and (4.2) immediately imply:

Lemma 4. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\mu_{1,C}^{FT}(s) < \mu_{2,C}^{FT}(s)$ and $\mu_{1,C}^{P}(s) < \mu_{2,C}^{P}(s)$, 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 $qk(s) < \mu_{1,C}^{FT}(s)$, we have $\tau = FT$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\mu_{1,C}^{FT}(s), \mu_{2,C}^{FT}(s)\right]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu_{2,C}^{FT}(s)$, we have $\tau = P$ and no dispute.*

2. *In states $s \in \sigma^{P}$: If DSB quality is high in the sense that $qk(s) < \mu_{1,C}^{P}(s)$, we have $\tau = P$ and no dispute; if DSB quality is intermediate in the sense that $qk(s) \in \left[\mu_{1,C}^{P}(s), \mu_{2,C}^{P}(s)\right]$, we have $\tau = P$ and a dispute; if DSB quality is low in the sense that $qk(s) > \mu_{2,C}^{P}(s)$, 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 $\sigma_{1,C}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) < \mu_{1,C}^{FT}(s)\right\}$, $\sigma_{2,C}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) \in \left[\mu_{1,C}^{FT}(s), \mu_{2,C}^{FT}(s)\right]\right\}$, and $\sigma_{3,C}^{FT} \equiv \left\{s \in \sigma^{FT} \mid qk(s) > \mu_{2,C}^{FT}(s)\right\}$, as well as $\sigma_{1,C}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) < \mu_{1,C}^{P}(s)\right\}$, $\sigma_{2,C}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) \in \left[\mu_{1,C}^{P}(s), \mu_{2,C}^{P}(s)\right]\right\}$, and, $\sigma_{3,C}^{P} \equiv \left\{s \in \sigma^{P} \mid qk(s) > \mu_{2,C}^{P}(s)\right\}$.

We can now characterize the efficiency loss associated with the V_C institution, $L(V_C)$, relative to the first-best outcome. There are two main changes relative to our baseline analysis (and thus two new terms relative to equation 3.3). First, if the court correctly rules FT there is now still a probability $m(s)$ that it awards excessively low damages d_{low}^{*} in which case Home will continue to choose $\tau = P$, leading to an additional efficiency loss (the first term on the last line in equation 4.3 below). Second, if the court incorrectly rules FT there is still a probability $1 - m(s)$ that it does not award excessively high damages in which case Home will continue

to choose $\tau = P$, leading to an additional efficiency gain (the second term on the last line in equation 4.3 below):

$$\begin{aligned} L(V_C) &= \sum_{s \in \sigma_{2,C}^{FT} \cup \sigma_{2,C}^{P}} p(s)\, qk(s)\, |\Gamma(s)| \\ &+ \sum_{s \in \sigma_{2,C}^{FT} \cup \sigma_{2,C}^{P}} p(s)\, [c(s) + c^*(s)] \\ &+ \sum_{s \in \sigma_{3,C}^{FT} \cup \sigma_{3,C}^{P}} p(s)\, |\Gamma(s)| \\ &+ \sum_{s \in \sigma_{2,C}^{FT}} p(s)\, [1 - qk(s)]\, m(s)\, |\Gamma(s)| - \sum_{s \in \sigma_{2,C}^{P}} p(s)\, qk(s)\, [1 - m(s)]\, |\Gamma(s)|. \end{aligned} \tag{4.3}$$

Finally, as noted above the equilibrium actions under the V_R institution are exactly the same as in the V_{G^*} institution of our baseline model, and hence 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 $\sigma_{2,R}^{FT} \equiv \sigma_{2,G^*}^{FT}$, $\sigma_{2,R}^{P} \equiv \sigma_{2,G^*}^{P}$, $\sigma_{3,R}^{FT} \equiv \sigma_{3,G^*}^{FT}$ and $\sigma_{3,R}^{P} \equiv \sigma_{3}^{P}$ where the latter sets are defined in our baseline model. With this we then have

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

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. Since the V_R institution mirrors the V_{G^*} institution from the baseline model, all action thresholds are also exactly the same and we now refer to them as $\mu_{1,R}^{FT}(s) \equiv \mu_1^{FT}(s)$, $\mu_{2,R}^{FT}(s) \equiv \mu_{2,G^*}^{FT}(s)$, $\mu_{1,R}^{P}(s) \equiv \mu_{1,G^*}^{P}(s)$, and $\mu_{2,R}^{P}(s) \equiv \mu_2^P(s)$ to aid comparison with the action thresholds associated with the V_C institution. As is easy to verify, the action thresholds associated with the V_C institution are below (above) the action thresholds associated with the V_R institution in states $s \in \sigma^{FT}$ (states $s \in \sigma^P$). Intuitively, the V_C institution gives the Home government the option to ignore the cease and desist order which makes

it more attractive for the Home government to protect and less attractive for the Foreign government to litigate. In states $s \in \sigma^{FT}$, the Home government therefore starts protecting earlier ($\mu_{1,C}^{FT}(s) < \mu_{1,R}^{FT}(s)$) and the Foreign government gives up litigating earlier ($\mu_{2,C}^{FT}(s) < \mu_{2,R}^{FT}(s)$) under the V_C institution; and in states $s \in \sigma^{P}$, the Foreign government starts litigating later ($\mu_{1,C}^{P}(s) > \mu_{1,R}^{P}(s)$) and the Home government gives up protecting later ($\mu_{2,C}^{P}(s) > \mu_{2,R}^{P}(s)$) under the V_C institution. Defining the sets $\Omega_{1,C}^{FT} \equiv \left\{ s \in \sigma^{FT} \mid qk(s) \in \left[\mu_{1,C}^{FT}(s), \mu_{1,R}^{FT}(s) \right] \right\}$, $\Omega_{2,C}^{FT} \equiv \left\{ s \in \sigma^{FT} \mid qk(s) \in \left[\mu_{2,C}^{FT}(s), \mu_{2,R}^{FT}(s) \right] \right\}$, $\Omega_{1,C}^{P} \equiv \left\{ s \in \sigma^{P} \mid qk(s) \in \left[\mu_{1,R}^{P}(s), \mu_{1,C}^{P}(s) \right] \right\}$, and $\Omega_{2,C}^{P} \equiv \left\{ s \in \sigma^{P} \mid qk(s) \in \left[\mu_{2,R}^{P}(s), \mu_{2,C}^{P}(s) \right] \right\}$, we can then write:

$$
\begin{aligned}
L(V_R) - L(V_C) \;=\; & -\sum_{s \in \sigma_{2,C}^{FT}} p(s)\left[1 - qk(s)\right] m(s) \left|\Gamma(s)\right| \\
& + \sum_{s \in \sigma_{2,C}^{P}} p(s)\, qk(s) \left[1 - m(s)\right] \left|\Gamma(s)\right| \\
& - \sum_{s \in \Omega_{1,C}^{FT}} p(s)\left[qk(s)\left|\Gamma(s)\right| + c(s) + c^{*}(s)\right] \\
& + \sum_{s \in \Omega_{1,C}^{P}} p(s)\left[qk(s)\left|\Gamma(s)\right| + c(s) + c^{*}(s)\right] \\
& + \sum_{s \in \Omega_{2,C}^{FT}} p(s)\left[qk(s)\left|\Gamma(s)\right| + c(s) + c^{*}(s) - \left|\Gamma(s)\right|\right] \\
& + \sum_{s \in \Omega_{2,C}^{P}} p(s)\left[\left|\Gamma(s)\right| - qk(s)\left|\Gamma(s)\right| - c(s) - c^{*}(s)\right]
\end{aligned}
\tag{4.4}
$$

Each line in this expression has an intuitive interpretation. The first line captures a disadvantage of the V_C institution, which is that correct court rulings are effectively overturned in states $s \in \sigma_{2,C}^{FT}$ if the court underestimates the damages. The second line then summarizes the corresponding advantage of the V_C institution, which is that incorrect court rulings are effectively overturned in states $s \in \sigma_{2,C}^{P}$ unless the court overestimates the damages. All other lines capture effects associated with movements in the action thresholds. In states $s \in \sigma^{FT}$, the inefficient policy is introduced earlier under the V_C institution, while in states $s \in \sigma^{P}$, the efficient policy is challenged later under the V_C institution, which is captured in the third and fourth line, respectively. In states $s \in \sigma^{FT}$, litigation against the inefficient policy is given up earlier under the V_C institution, while in states $s \in \sigma^{P}$, the efficient policy is given up later under the V_C institution, which is captured in the fifth and sixth line, respectively.

Two observations allow us to describe intuitive conditions under which expression (4.4) can

be signed. The first observation is that an improvement in the court's ability to assess damages tends to favor the V_C institution. To see this, notice that the sets $\Omega_{1,C}^{FT}$, $\Omega_{2,C}^{FT}$, and $\Omega_{1,C}^{P}$ become empty as $m(s) \to 0$ so that expression (4.4) reduces to:

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

Correctly assessed cash damages make the Home government fully internalize the effect its trade policy choices have on the Foreign government, so that it stops protecting when (rightly) convicted in states $s \in \sigma_{2,C}^{FT}$ and continues protecting when (wrongly) convicted in states $s \in \sigma_{2,C}^{P}$.[30] As a result, there are no longer any inefficient breaches of the court's cease and desist order, and the welfare gains associated with the efficient breaches are captured in the first line of the expression above. However, the V_C institution also brings about additional litigation in states $s \in \Omega_{2,C}^{P}$, which could be socially desirable or not. This is captured in the second line of the above expression, which compares the welfare losses associated with the V_R institution in states $s \in \Omega_{2,C}^{P}$ ($|\Gamma(s)|$) to the expected welfare losses associated with the V_C institution in these states ($qk(s)\,|\Gamma(s)| + c(s) + c^{*}(s)$). Notice that the second line is positive if the joint litigation costs $c(s) + c^{*}(s)$ are sufficiently small. Overall, the V_C institution therefore dominates the V_R institution if the court's ability to assess damages is sufficiently good, assuming that the joint litigation costs are sufficiently small.

The second observation is that an increase in the probability that free trade is the efficient policy tends to favor the V_R institution. To see this, notice that the sets $\sigma_{2,C}^{P}$, $\Omega_{1,C}^{P}$, and $\Omega_{2,C}^{P}$ become empty as $p^{FT} \to 1$ so that expression (4.4) reduces to:

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

[30] 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 $s \in \sigma_{3,C}^{P}$ as well as $\tau = P$ in states $s \in \sigma_{3,C}^{FT}$, where court quality is so bad that the court is not even invoked.

If free trade is the efficient policy, allowing the Home government to ignore the court's cease and desist ruling by paying cash damages brings about a welfare loss, and this welfare loss is captured in the first line of the expression above. The other two lines summarize the welfare effects associated with the more aggressive protectionism of the Home government in states $s \in \Omega_{1,C}^{FT}$ and $s \in \Omega_{2,C}^{FT}$. While the second term is unambiguously negative (the efficient policy gets revoked earlier under the cash institution), the third term can be positive or negative (the additional litigation under the cash institution may be socially desirable or not), and becomes negative if the joint litigation costs $c(s)+c^*(s)$ are sufficiently small. Overall, the V_R institution therefore dominates the V_C institution if free trade is sufficiently likely to be the efficient trade policy, assuming that the joint litigation costs are sufficiently small.

Imposing that the joint litigation costs are sufficiently small, we can therefore summarize:

Proposition 4. *Allowing for retaliation instead of cash damages in a trade agreement is optimal (1) if the court's ability to assess cash damages is sufficiently bad, 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 investment treaties. As indicated above, we assume that foreign investors have standing in investment disputes, and we again take retaliation to be a sufficiently inefficient form of damage payment that under retaliation our earlier analysis of investment treaties with ISDS still applies. Our task is then to introduce cash damages into our model of investment treaties, and compare the outcomes from our earlier model to the outcomes under cash damages derived here.

As before, we assume that the court-assessed damages are realizations of a random variable. Denoting the damages awarded to the injured foreign investor by $\bar{d}^*(I^*)$, we assume $\Pr\left[\bar{d}^*(I^*) = \bar{\gamma}_{I^*}^*(I^*)\right] = 1-2m(s)$, $\Pr\left[\bar{d}^*(I^*) > \bar{\gamma}_{I^*}^*(I^*)\right] = m(s)$, and $\Pr\left[\bar{d}^*(I^*) < \bar{\gamma}_{I^*}^*(I^*)\right] = m(s)$, with $m(s) \in [0, \frac{1}{2}]$. As before, the court awards the correct damages with probability $1-2m(s)$ and overestimates or underestimates the damages with symmetric probabilities $m(s)$. Defining $\bar{d}_{high}^*(I^*) \equiv E\left[\bar{d}^*(I^*) \mid \bar{d}^*(I^*) > \bar{\gamma}_{I^*}^*(I^*)\right]$ and $\bar{d}_{low}^* \equiv E\left[\bar{d}^*(I^*) \mid \bar{d}^*(I^*) > \bar{\gamma}_{I^*}^*(I^*)\right]$, we further assume that $\bar{d}_{high}^*(I^*) > \bar{\gamma}_G(I^*, s)$ and $\bar{d}_{low}^* < \bar{\gamma}_G(I^*, s)$ for all s. So as before, if the court overestimates the damages, Home can expect to be assessed damages in the amount of $\bar{d}_{high}^*$, which would be higher than Home's valuation of the violation $\bar{\gamma}_G(I^*, s)$. Conversely, if the court underestimates the damages, Home can expect to be assessed damages in the amount

of $\bar{d}^*_{low}$ which would be lower than its valuation of the violation $\bar{\gamma}_G(I^*, s)$. Again 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

To assess whether there are conditions under which cash damages are optimal in an investment treaty, 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

$$\Pr(\text{ruling is } FT \mid s)\left\{\Pr(\iota = FT)\left|\bar{\gamma}^*_{I^*}(I^*)\right| + \Pr(\iota = T)\, E\left[\bar{d}^*(I^*) \mid \iota = T\right]\right\} > \bar{c}^*(I^*, s). \tag{4.5}$$

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

$$\begin{aligned}&\Pr(\text{ruling is } T \mid s)\,\bar{\gamma}_G(I^*, s) + \\ &\Pr(\text{ruling is } FT \mid s)\Pr(\iota = T)\left\{\bar{\gamma}_G(I^*, s) - E\left[\bar{d}^*(I^*) \mid \iota = T\right]\right\} > \bar{c}(I^*, s).\end{aligned} \tag{4.6}$$

Analogous to before, here $\Pr(\iota = FT)$ is the probability that the Host government chooses to cease and desist in the face of a court ruling of FT while $\Pr(\iota = T)$ is the probability that the Host government instead continues with $\iota = T$ and pays damages, so that $\Pr(\iota = FT) = 1 - \Pr(\iota = T)$.

We can now derive the equilibrium actions for each state s. Defining the thresholds $\bar{\mu}^{FT}_{1,C}(s) \equiv \frac{\bar{c}\epsilon(s)\bar{\gamma}_G(I^*,s) - m(s)\left[\bar{\gamma}_G(I^*,s) - \bar{d}^*_{low}\right]}{\bar{\gamma}_G(I^*,s) - m(s)\left[\bar{\gamma}_G(I^*,s) - \bar{d}^*_{low}\right]}$, $\bar{\mu}^{FT}_{2,C}(s) \equiv 1 - \frac{\bar{c}^*\epsilon^*(s)PS(I^*)}{[1-2m(s)]PS(I^*) + m(s)\bar{d}^*_{low}}$, $\bar{\mu}^{T}_{1,C}(s) \equiv \frac{\bar{c}^*\epsilon^*(s)PS(I^*)}{m(s)PS(I^*) + [1-2m(s)]PS(I^*) + m(s)\bar{d}^*_{low}}$, $\bar{\mu}^{T}_{2,C}(s) \equiv \frac{\bar{\gamma}_G(I^*,s) - \bar{c}\epsilon(s)\bar{\gamma}_G(I^*,s)}{\bar{\gamma}_G(I^*,s) - \left\{[1-2m(s)][\bar{\gamma}_G(I^*,s) - PS(I^*)] + m(s)\left[\bar{\gamma}_G(I^*,s) - \bar{d}^*_{low}\right]\right\}}$, and noting as before that $\bar{\mu}^{FT}_{1,C}(s) < \bar{\mu}^{FT}_{2,C}(s)$ and $\bar{\mu}^{P}_{1,C}(s) < \bar{\mu}^{P}_{2,C}(s)$ if the dispute costs are low relative to the dispute stakes for each disputant, conditions (4.5) and (4.6) immediately imply the following result:

Lemma 5. *Assuming that dispute costs are low relative to dispute stakes for all s so that $\bar{\mu}^{FT}_{1,C}(s) < \bar{\mu}^{FT}_{2,C}(s)$ and $\bar{\mu}^{P}_{1,C}(s) < \bar{\mu}^{P}_{2,C}(s)$, 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* $qk(s) < \bar{\mu}^{FT}_{1,C}(s)$, *we have* $\iota = FT$ *and no dispute; if DSB quality is intermediate in the sense that* $qk(s) \in \left[\bar{\mu}^{FT}_{1,C}(s), \bar{\mu}^{FT}_{2,C}(s)\right]$, *we have* $\iota = T$ *and a dispute; if DSB quality is low in the sense that* $qk(s) > \bar{\mu}^{FT}_{2,C}(s)$, *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* $qk(s) < \bar{\mu}^{T}_{1,C}(s)$, *and a dispute if* $qk(s) > \bar{\mu}^{T}_{1,C}(s)$.
2. *If the Host government is unconstrained: If DSB quality is high in the sense that* $qk(s) < \bar{\mu}^{T}_{1,C}(s)$, *we have* $\iota = T$ *and no dispute; if DSB quality is intermediate in the sense that* $qk(s) \in \left[\bar{\mu}^{T}_{1,C}(s), \bar{\mu}^{T}_{2,C}(s)\right]$, *we have* $\iota = T$ *and a dispute; if DSB quality is low in the sense that* $qk(s) > \bar{\mu}^{T}_{2,C}(s)$, *we have* $\iota = FT$ *and no dispute.*

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

With this lemma, we now characterize the expected welfare of the Host government under an investment treaty with cash damages, $E_s\left[\bar{\omega}^{C}_{I^*}(I^*, s)\right]$, using the fact that the expected return to the Foreign investor on an investment of I^* under cash damages must equal the world interest

rate r^* net of the investment incentive. As we show in the Appendix, this yields:

$$
\begin{aligned}
& E_s\left[\bar{\omega}_{I^*}^{C}\left(I^*,s\right)\right] \qquad (4.7)\\
= \; & (1-\bar{p})\sum_{s\in\bar{\sigma}_{1,C}^{FT}} p\left(s\right)\left[CS\left(I^*\right)+PS\left(I^*\right)\right]\\
& +(1-\bar{p})\sum_{s\in\bar{\sigma}_{2,C}^{FT}} p\left(s\right)\left\{CS\left(I^*\right)+PS\left(I^*\right)-\left\{qk\left(s\right)+\left[1-qk\left(s\right)\right]m\left(s\right)\right\}\left(1-\kappa\right)PS\left(I^*\right)-\bar{c}\left(I^*,s\right)-\bar{c}^*\left(I^*,s\right)\right\}\\
& +(1-\bar{p})\sum_{s\in\bar{\sigma}_{3,C}^{FT}} p\left(s\right)\left[CS\left(I^*\right)+\kappa PS\left(I^*\right)\right]\\
& -(1-\bar{p})\sum_{s\in\bar{\sigma}_{2,C}^{T}} p\left(s\right)\left\{qk\left(s\right)m\left(s\right)\left[e\left(I^*,s\right)-CS\left(I^*\right)-PS\left(I^*\right)\right]+\bar{c}\left(I^*,s\right)+\bar{c}^*\left(I^*,s\right)\right\}\\
& -(1-\bar{p})\sum_{s\in\bar{\sigma}_{3,C}^{T}} p\left(s\right)\left[e\left(I^*,s\right)-CS\left(I^*\right)-PS\left(I^*\right)\right]\\
& +\bar{p}p^{FT}\left[CS\left(I^*\right)+PS\left(I^*\right)\right]\\
& -\bar{p}\sum_{s\in\bar{\sigma}_{2,C}^{T}\cup\bar{\sigma}_{3,C}^{T}} p\left(s\right)\left[\bar{c}\left(I^*,s\right)+\bar{c}^*\left(I^*,s\right)\right]\\
& -r^*I^*.
\end{aligned}
$$

There are three substantive differences between this expression and its analog from the baseline model (equation 3.14): First, in states $s\in\bar{\sigma}_{2,C}^{FT}$ there is now a higher probability of losing $\left(1-\kappa\right)PS\left(I^*\right)$ in social surplus since the Host government now chooses $\iota=T$ even if it is rightly convicted by the DSB if the DSB awards excessively low damages (the third line of both expressions). Second, in states $s\in\bar{\sigma}_{2,C}^{T}$, there is now a lower probability of losing $e\left(I^*,s\right)-CS\left(I^*\right)-PS\left(I^*\right)$ in social surplus since the Host government only follows an incorrect cease and desist order if the assessed damage payments are excessively high (the fourth line of both expressions). And finally, in states $s\in\bar{\sigma}_{2,C}^{T}\cup\bar{\sigma}_{3,C}^{T}$ there is now a zero probability of losing $e\left(I^*,s\right)-CS\left(I^*\right)-PS\left(I^*\right)$ in social surplus if the Host government is constrained in its policy choices simply because it is now possible to continue implementing the first-best policy in case of an erroneous conviction (the seventh line of both expressions).

4.2.1. Cash versus Retaliation

We now provide sufficient conditions under which the Host government's ex-ante expected welfare under an investment treaty with cash damages exceeds its ex-ante expected welfare under an investment treaty that relies on retaliation, with the latter given by (3.14) and which we now de-

note by $E_s\left[\bar{\omega}^R_{I^*}\left(\bar{I}^{*R},s\right)\right]$. Noting that $\bar{\mu}^{FT}_{1,C}(s)<\bar{\mu}^{FT}_{1,R}(s)$, $\bar{\mu}^{FT}_{2,C}(s)<\bar{\mu}^{FT}_{2,R}(s)$, $\bar{\mu}^{T}_{1,C}(s)>\bar{\mu}^{T}_{1,R}(s)$, and $\bar{\mu}^{T}_{2,C}(s)>\bar{\mu}^{T}_{2,R}(s)$, it follows that we can again define sets $\bar{\Omega}^{FT}_{1,C}$, $\bar{\Omega}^{FT}_{2,C}$, $\bar{\Omega}^{T}_{1,C}$, and $\bar{\Omega}^{T}_{2,C}$, using the same notation convention as above. Denoting by $\bar{I}^{*R}$ the optimal level of investment under an investment treaty with retaliation and an ISDS as characterized in (3.15), we may then use (4.7) to write

$$
\begin{aligned}
& E_s\left[\bar{\omega}^C_{I^*}\left(\bar{I}^{*R},s\right)\right]-E_s\left[\bar{\omega}^R_{I^*}\left(\bar{I}^{*R},s\right)\right] \qquad (4.8)\\
= \;& -(1-\bar{p})\sum_{s\in\bar{\sigma}^{FT}_{2,C}} p(s)\left[1-qk(s)\right]m(s)(1-\kappa)PS\left(\bar{I}^{*R}\right)\\
& +(1-\bar{p})\sum_{s\in\bar{\sigma}^{T}_{2,C}} p(s)\,qk(s)\left[1-m(s)\right]\left[e\left(\bar{I}^{*R},s\right)-CS\left(\bar{I}^{*R}\right)-PS\left(\bar{I}^{*R}\right)\right]\\
& -(1-\bar{p})\sum_{s\in\bar{\Omega}^{FT}_{1,C}} p(s)\left[qk(s)(1-\kappa)PS\left(\bar{I}^{*R}\right)+\bar{c}\left(\bar{I}^{*R},s\right)+\bar{c}^*\left(\bar{I}^{*R},s\right)\right]\\
& +(1-\bar{p})\sum_{s\in\bar{\Omega}^{T}_{1,C}} p(s)\left\{qk(s)\left[e\left(\bar{I}^{*R},s\right)-CS\left(\bar{I}^{*R}\right)-PS\left(\bar{I}^{*R}\right)\right]+\bar{c}\left(\bar{I}^{*R},s\right)+\bar{c}^*\left(\bar{I}^{*R},s\right)\right\}\\
& -(1-\bar{p})\sum_{s\in\bar{\Omega}^{FT}_{2,C}} p(s)\left\{\left[1-qk(s)\right](1-\kappa)PS\left(\bar{I}^{*R}\right)-\bar{c}\left(\bar{I}^{*R},s\right)-\bar{c}^*\left(\bar{I}^{*R},s\right)\right\}\\
& +(1-\bar{p})\sum_{s\in\bar{\Omega}^{T}_{2,C}} p(s)\left\{\left[1-qk(s)\right]\left[e\left(\bar{I}^{*R},s\right)-CS\left(\bar{I}^{*R}\right)-PS\left(\bar{I}^{*R}\right)\right]-\bar{c}\left(\bar{I}^{*R},s\right)-\bar{c}^*\left(\bar{I}^{*R},s\right)\right\}\\
& +\bar{p}\sum_{s\in\bar{\Omega}^{T}_{1,C}} p(s)\left[\bar{c}\left(\bar{I}^{*R},s\right)+\bar{c}^*\left(\bar{I}^{*R},s\right)\right]\\
& +\bar{p}\sum_{s\in\bar{\sigma}^{T}_{2,R},\bar{\sigma}^{T}_{3,R}} p(s)\,qk(s)\left[e\left(\bar{I}^{*R},s\right)-CS\left(\bar{I}^{*R}\right)-PS\left(\bar{I}^{*R}\right)\right].
\end{aligned}
$$

Clearly, if $E_s\left[\bar{\omega}^C_{I^*}\left(\bar{I}^{*R},s\right)\right]-E_s\left[\bar{\omega}^R_{I^*}\left(\bar{I}^{*R},s\right)\right]\geq 0$, it must also be that $E_s\left[\bar{\omega}^C_{I^*}\left(\bar{I}^{*C},s\right)\right]-E_s\left[\bar{\omega}^R_{I^*}\left(\bar{I}^{*R},s\right)\right]\geq 0$, where $\bar{I}^{*C}$ is the level of investment optimal under an investment treaty with cash damages and an ISDS. For this reason (4.8) allows us to establish sufficient conditions under which an investment treaty with cash damages is preferred to an investment treaty with retaliation, similar to how our earlier equation (4.4) allowed us to establish conditions under which a trade agreement with retaliation is preferred to a trade agreement with cash damages. The individual lines of both equations have a very similar interpretation, keeping in mind of course that the Host government is now constrained to choose the first-best policies with probability $\bar{p}$ (the last two lines).

Just as we saw in the case of a trade agreement, an improvement in the court's ability to

assess damages tends to favor the cash institution also in the case of investment agreements. To see this, notice that the sets $\bar{\Omega}_{1,C}^{FT}$, $\bar{\Omega}_{2,C}^{FT}$, and $\bar{\Omega}_{1,C}^{T}$ become empty as $m(s) \to 0$ so that equation (4.8) becomes:

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

The term on the second line is unambiguously positive (capturing the benefits of efficient breach under cash damages), while the term on the third line can be positive or negative in principle but becomes positive if the joint litigation costs are sufficiently small (capturing the social net benefits of the additional litigation under cash damages). The term on the fourth line is new relative to our earlier discussion of trade agreements and refers to the case in which the Host government can commit to implementing the first-best policies. It is unambiguously positive since the Host government is always able to implement the first best policy under cash damages by paying damages even if the court makes a mistake. Overall, an investment treaty with cash damages therefore dominates an investment treaty with retaliation if the court's ability to assess damages is sufficiently good, assuming that the joint litigation costs are sufficiently small.

Also just as in the case of a trade agreement, an increase in the probability that free trade is the efficient policy tends to favor an investment treaty with retaliation. To see this, notice that the sets $\sigma_{2,C}^{T}$, $\Omega_{1,C}^{T}$, and $\Omega_{2,C}^{T}$ become empty as $p^{FT} \to 1$ so that expression (4.8) reduces to:

$$\begin{aligned}
& E_s\left[\bar{\omega}_{I^*}^{C}\left(\bar{I}^{*R}, s\right)\right] - E_s\left[\bar{\omega}_{I^*}^{R}\left(\bar{I}^{*R}, s\right)\right] \\
= \; & -(1-\bar{p}) \sum_{s \in \bar{\sigma}_{2,C}^{FT}} p(s)\,[1-qk(s)]\, m(s)\,(1-\kappa)\, PS\left(\bar{I}^{*R}\right) \\
& -(1-\bar{p}) \sum_{s \in \bar{\Omega}_{1,C}^{FT}} p(s)\left[qk(s)\,(1-\kappa)\, PS\left(\bar{I}^{*R}\right) + \bar{c}\left(\bar{I}^{*R}, s\right) + \bar{c}^{*}\left(\bar{I}^{*R}, s\right)\right] \\
& -(1-\bar{p}) \sum_{s \in \bar{\Omega}_{2,C}^{FT}} p(s)\left\{[1-qk(s)]\,(1-\kappa)\, PS\left(\bar{I}^{*R}\right) - \bar{c}\left(\bar{I}^{*R}, s\right) - \bar{c}^{*}\left(\bar{I}^{*R}, s\right)\right\}
\end{aligned}$$

The terms on the second and third lines are unambiguously negative (capturing the costs of inefficient breach and more aggressive protectionism under cash damages, respectively). The term on the last line can be positive or negative in principle but is negative if the joint litigation costs are sufficiently small (capturing the social net benefits of the additional litigation under cash damages in states $s \in \bar{\Omega}_{2,C}^{FT}$). Overall, an investment treaty with retaliation therefore dominates an investment treaty with cash damages if free trade is sufficiently likely to be the efficient investment policy, assuming that the joint litigation costs are sufficiently small.

Imposing that the joint litigation costs are sufficiently small, we can therefore summarize:

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

In effect, Propositions 4 and 5 imply that it makes sense that investment treaties 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. In practice, this condition seems plausible at least at a broad level, in light of the difficulty of placing a monetary value on the kinds of issues, such as lost jobs and distributional considerations, that typically dominate the trade concerns of real-world governments, and the relative ease by comparison of assessing lost profits from a (regulatory or outright) taking.

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: Can our models make sense of why trade agreements adopt prospective remedies while investment treaties adopt retrospective remedies? That is, can our models help us understand the conditions under which it might be optimal for investment treaties to include retrospective damages in the remedy (damages for harm suffered before the case is finally adjudicated) and

for trade agreements to limit the remedy to only prospective damages (damages that would arise after adjudication if the ruling is not obeyed)?

To capture these alternative remedial periods, 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 investment treaties 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.

Building on our analysis in section 3, we assume that the trade agreement has adopted SSDS while the investment treaty has adopted ISDS. And building on our analysis of section 4, for the retrospective damages we assume that the trade agreement relies on retaliation for damage payments while the investment treaty employs cash, though in contrast to 4 we now assume that the court can perfectly assess the level of damages so that we can focus on the inefficiency of retaliation relative to cash as a form of damage payments.[31] And finally, to keep the comparison clean we continue to assume that the prospective remedy for both trade agreements and investment treaties is a cease and desist order, just as in the models of section 3: this means that 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, a feature that makes our comparisons easier but is not necessary for our results.

[31] As we note, our assumption that for the retrospective damages the trade agreement relies on retaliation while the BIT relies on cash 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).

5.1. Trade Agreements

We consider first the case of a trade agreement, and look for conditions under which prospective remedies would be optimal. To this end, we consider the implications of adopting retrospective remedies in a trade agreement. Under retrospective remedies, 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).[32]

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 (who would reduce these other tariffs) or the exporter government (who would raise tariffs and hence engage in reciprocal retaliation), that amount to a *costly transfer* to the exporter government.[33] 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.

Adopting retrospective remedies will have important implications for the conditions describing equilibrium behavior of the two governments. Consider first the foreign government's filing behavior under a retrospective remedy. 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

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

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 a retrospective remedy. This

[32] 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.

[33] 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.

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:

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

or

$$\Pr(\text{DSB ruling is } P \mid s) \times \gamma_G(s) + \Pr(\text{DSB ruling is } FT \mid s) \times \{\delta[\gamma_G(s) - \frac{|\gamma^*_{G^*}(s)|}{\beta}]\} > c(s). \tag{5.2}$$

The important novel element of (5.2) is that if the Home government chooses P and the DSB rules in favor of FT, the Home government will be responsible for compensating the foreign government for the harm done prior to the ruling (with the retroactive damage payment $\delta\frac{|\gamma^*_{G^*}(s)|}{\beta}$), and there is no action that the Home government can take to avoid making these damage payments once they are assessed (i.e., while the harm to the foreign government going forward can be removed by reverting to a policy of FT, this does nothing to address the retrospective harm).

We can now derive the equilibrium actions for each state s in the presence of retrospective remedies. As above 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 $\mu^{FT}_{1,Retro}(s) \equiv \frac{c(s)-\delta\gamma_G(s)+\delta\frac{|\gamma^*_{G^*}(s)|}{\beta}}{[1-\delta]\gamma_G(s)+\delta\frac{|\gamma^*_{G^*}(s)|}{\beta}}$, $\mu^{FT}_{2,Retro}(s) \equiv 1 - \frac{c^*(s)}{|\gamma^*_{G^*}(s)|}$, $\mu^{P}_{1,Retro}(s) \equiv \frac{c^*(s)}{|\gamma^*_{G^*}(s)|}$, 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 (5.1) 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 $\mu^{FT}_{1,Retro}(s) < \mu^{FT}_{2,Retro}(s)$ 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.*

Lemma 6 reveals a key point: for any $\delta > 0$ and as β approaches zero, $\mu^{FT}_{1,Retro}(s)$ approaches one and $\mu^P_{2,Retro}(s)$ approaches zero, 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 $s \in \sigma^P$).[34] Put differently, for any fixed court quality q, the joint surplus under a trade agreement with retrospective remedies will approach the joint surplus associated with FT in all states as β approaches zero and the cost of transfers becomes prohibitive. We record this in:

Remark 1. *If transfers in the context of a trade dispute are sufficiently costly (β small), then for any $\delta > 0$ the joint surplus under a trade agreement with retrospective remedies will approach the joint surplus associated with FT in all states, no matter how accurate the court may be (for any $q > 0$).*

Remark 1 implies that for any $\delta > 0$ and $q > 0$, if β is sufficiently small then a trade agreement with retrospective remedies will be dominated by a trade agreement with prospective remedies provided that the trade agreement with prospective remedies delivers a level of joint surplus higher than that associated with FT in all states. But if the quality of the court q is fixed at a sufficiently high level, then for any level of β as δ approaches zero so that litigation delay becomes sufficiently short, the joint surplus under a trade agreement with prospective remedies 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 trade agreement with retrospective remedies under these conditions.

We may therefore state:

Proposition 6. *A prospective remedy is optimal for a trade agreement 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).*

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

5.2. Investment Agreements

We now turn to the case of an investment treaty, and look for conditions under which retrospective remedies would be optimal. To this end, we consider the implications of adopting prospective remedies in an investment treaty. Relative to our earlier analysis of investment treaties (i.e., relative to the case where $\delta = 0$), for $\delta > 0$ the conditions describing equilibrium behavior of the Host government and the foreign investor are altered.

Consider first the foreign investor's filing behavior under a prospective remedy. 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

$$\Pr(\text{DSB ruling is } FT \mid s) \times [1-\delta]|\bar{\gamma}^*_{I^*}(I^*)| > \bar{c}^*(I^*, s). \tag{5.3}$$

Condition (5.3) 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 (5.3) fails – because then the Host government can set $\iota = T$ without triggering a dispute – or if (5.3) holds and the expected benefit to the Host government from a taking exceeds the cost to the Host government of a dispute:

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

or

$$\Pr(\text{DSB ruling is } FT \mid s) \times \delta\bar{\gamma}_G(I^*, s) + \Pr(\text{DSB ruling is } T \mid s) \times \bar{\gamma}_G(I^*, s) > \bar{c}(I^*, s). \tag{5.4}$$

The important novel element of (5.4) is that if the Host government chooses T and the DSB rules in favor of FT, the Host government still enjoys the benefits of T for the pre-ruling period $(\delta\bar{\gamma}_G(I^*, s))$.

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 $\bar{\mu}^{FT}_{1,Pro}(s) \equiv \frac{\bar{c}\epsilon(s)-\delta}{[1-\delta]}$, $\bar{\mu}^{FT}_{2,Pro}(s) \equiv 1 - \frac{\bar{c}^*\epsilon^*(s)}{[1-\delta]}$, $\bar{\mu}^{T}_{1,Pro}(s) \equiv \frac{\bar{c}^*\epsilon^*(s)}{[1-\delta]}$, and

$\bar{\mu}^{T}_{2,Pro}(s) \equiv 1 - \frac{\bar{c}\epsilon(s)-\delta}{[1-\delta]}$, 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 (5.3) and (5.4) immediately imply the following result:

Lemma 7. *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 investment treaties under ISDS.[35] On the other hand, if δ is sufficiently close to one, (5.3) and (5.4) 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 investment treaty with prospective remedies would be 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,

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

the critical level of δ beyond which an investment treaty with prospective remedies would be valueless, which we denote by $\bar{\delta}$, is defined by

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

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). We record this in:

Remark 2. *If litigation delay is sufficiently high (for $\delta \geq \bar{\delta}$), the Host government cannot improve upon a stand-alone program of up-front investment incentives to foreign investors by introducing an investment treaty with prospective remedies, no matter how accurate the court may be (for any $q > 0$).*

Remark 2 implies that for any $q > 0$, if δ is sufficiently high then an investment treaty with prospective remedies will be dominated by an investment treaty with retrospective remedies provided that the investment treaty with retrospective remedies can improve upon a stand-alone program of up-front investment incentives to foreign investors. But if the quality of the court q is fixed at a sufficiently high level, then for any level of δ as β approaches one so that transfers in the context of an investment treaty are sufficiently efficient, the Host government surplus under an investment treaty with retrospective remedies can be brought arbitrarily close to the first best level, which exceeds the Host government surplus under a stand-alone program of up-front investment incentives and therefore beats an investment treaty with prospective remedies under these conditions.

We may therefore state:

Proposition 7. *A retrospective remedy is optimal for an investment treaty provided that the degree of litigation delay is sufficiently long (δ large), transfers in the context of an investment treaty are sufficiently efficient (β large) and the quality of the court is sufficiently high (q low).*

Summarizing, Propositions 6 and 7 imply the following. If court quality is sufficiently high (q is sufficiently low), it is optimal to adopt prospective remedies in trade agreements such as the WTO while it is optimal to adopt retrospective remedies in investment treaties

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 remedies 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 remedies 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.

We have identified conditions under which, at least at a broad level, the observed 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.

These conditions strike us as plausible, though far from universal, features of the trade and investment dispute settlement environments that are our focus. In this light, our results can help interpret the observed 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 foreign investors to bring claims against

governments in investment disputes, are not optimal under all circumstances. For example, according to our results, allowing foreign investors to have standing in an investment treaty may not be optimal for a government that is attempting to make commitments to those investors, if the investors' own governments bear low political costs from the initiation of disputes on their investors' behalf. Likewise, if host governments have found other means to make commitments to investors and investment treaties are primarily concerned with market access rather than commitment issues, our results would point to a strong presumption against providing standing to foreign investors. This all suggests that, at a minimum, such features deserve closer scrutiny before drawing the conclusion that existing treaty arrangements are necessarily optimal.

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 (2018). "Investor-State vs. State-State Dispute Settlement," IFN WP No 1248.

Horn, Henrik and Thomas Tangeras (2019). "Economics and Politics of International Investment Agreements," Research Institute of Industrial Economics, IFN WP No. 1140, revised.

Janeba, Eckhard, (2019). "Regulatory Chill and the Effects of Investor State Dispute Settlements," *Review of International Economics* 27: 1172-1198.

Kohler, Wilhelm and Frank Stahler (2019). "The Economics of Investor Protection: ISDS versus National Treatment," Journal of International Economics 121.

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 (in progress). "FDI, Exports and the Terms-of-Trade Motive for Tariff Intervention," April.

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 WP, No. 235.

Sykes, Alan O. (2019). "The Economic Structure of International Investment Agreements with Implications for Treaty Interpretation and Design," *American Journal of International Law* 113: 482-534.

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. Appendix

7.1. Derivation of equation (3.14)

Given the equilibrium actions described by Lemma 3, the expected return on a level of FDI I^* from expected ex-post operating profits is given by:

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

Keeping in mind that the foreign investor must be offered an up-front investment incentive $\left\{r^*-E_s\left[\rho_f^*\left(I^*,s\right)\right]\right\}I^*$ in order to be willing to invest I^* given its outside option, the Host government's expected payoff can be written as:

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

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 equation (3.14) in the main text.

7.2. Derivation of equation (4.7)

Given the equilibrium actions described by Lemma 5, the expected return on a level of FDI I^* from expected ex-post operating profits is given by:

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

Keeping in mind that the foreign investor must be offered an up-front investment incentive $\left\{r^*-E_s\left[\rho_{I^*}^{*C}\left(I^*,s\right)\right]\right\}I^*$ in order to be willing to invest I^* given its outside option, the Host government's expected payoff can be written as:

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

Plugging the expression for $E_s\left[\rho_{I^*}^{*C}\left(I^*,s\right)\right]$ into the above expression for $E_s\left[\omega_{I^*}^{C}\left(I^*,s\right)\right]$ yields equation (4.7) in the main text.

7.3. Analysis of optimal investment incentives under cash institution

The optimal investment incentives under the cash institution can be found by maximizing (4.7) with respect to I^*. The associated first-order condition can be manipulated to yield the following implicit characterization of the optimal investment level $\bar{I}_C^*$:

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

This expression is the analog to (3.15) in the main text. The second line is different because the Host government now also chooses the inefficient policy $\iota = T$ in states $s \in \sigma^{FT}$ if the court rules correctly but underestimates the damages. The fourth line is different because the Host government now continues to choose the efficient policy $\iota = T$ in states $s \in \sigma^{T}$ even if the court rules incorrectly unless the court also overestimates the damages. And there is no sixth line because the Host government now implements the efficient policy regardless of the court ruling (potentially paying the awarded damages of course).

Plugging this expression into (4.7) and simplifying yields:

$$
\begin{aligned}
& E_s\left[\omega_{I^*}^C\left(\bar{I}_C^*, s\right)\right] \\
= \; & (1-\bar{p})\left[p^{FT}+p_{3,C}^T+\sum_{s\in\bar{\sigma}_{2,C}^T} p(s)\, qk(s)\, m(s)\right] CS\left(\bar{I}_C^*\right) \\
& +(1-\bar{p})(1-\kappa)\left\{\sum_{s\in\bar{\sigma}_{2,C}^{FT}} p(s)\left\{qk(s)+[1-qk(s)]\, m(s)\right\}+p_{3,C}^{FT}\right\}\left(\bar{I}_C^*\right)^2 \frac{\partial\tilde{P}\left(\bar{I}_C^*\right)}{\partial I^*} \\
& -(1-\bar{p})\sum_{s\in\bar{\sigma}_{2,C}^{FT}\cup\bar{\sigma}_{2,C}^T} p(s)\left\{c\left(\bar{I}_C^*, s\right)+c^*\left(\bar{I}_C^*, s\right)-\bar{I}_C^*\left[\frac{\partial c\left(\bar{I}_C^*, s\right)}{\partial I^*}+\frac{\partial c^*\left(\bar{I}_C^*, s\right)}{\partial I^*}\right]\right\} \\
& +\bar{p}p^{FT} CS\left(\bar{I}_C^*\right) \\
& -\bar{p}\sum_{s\in\bar{\sigma}_{2,C}^T\cup\bar{\sigma}_{3,C}^T} p(s)\left\{c\left(\bar{I}_C^*, s\right)+c^*\left(\bar{I}_C^*, s\right)-\bar{I}_C^*\left[\frac{\partial c\left(\bar{I}_C^*, s\right)}{\partial I^*}+\frac{\partial c^*\left(\bar{I}_C^*, s\right)}{\partial I^*}\right]\right\}
\end{aligned}
$$

This expression is the analog to (3.16) in the main text. The second line is different because the Host government now chooses the inefficient policy $\iota = T$ in states $s \in \sigma^{FT}$ even when correctly convicted as long as the court underestimates the damages. The fourth line is also different since the Host government now implements the first-best policy no matter what when it is constrained by domestic institution. Using this expression, it is easy to verify following a similar logic as in the baseline model that the investment treaty with cash damages only improves upon the program of up-front investment incentives if court quality is sufficiently high and domestic institutions are sufficiently weak.

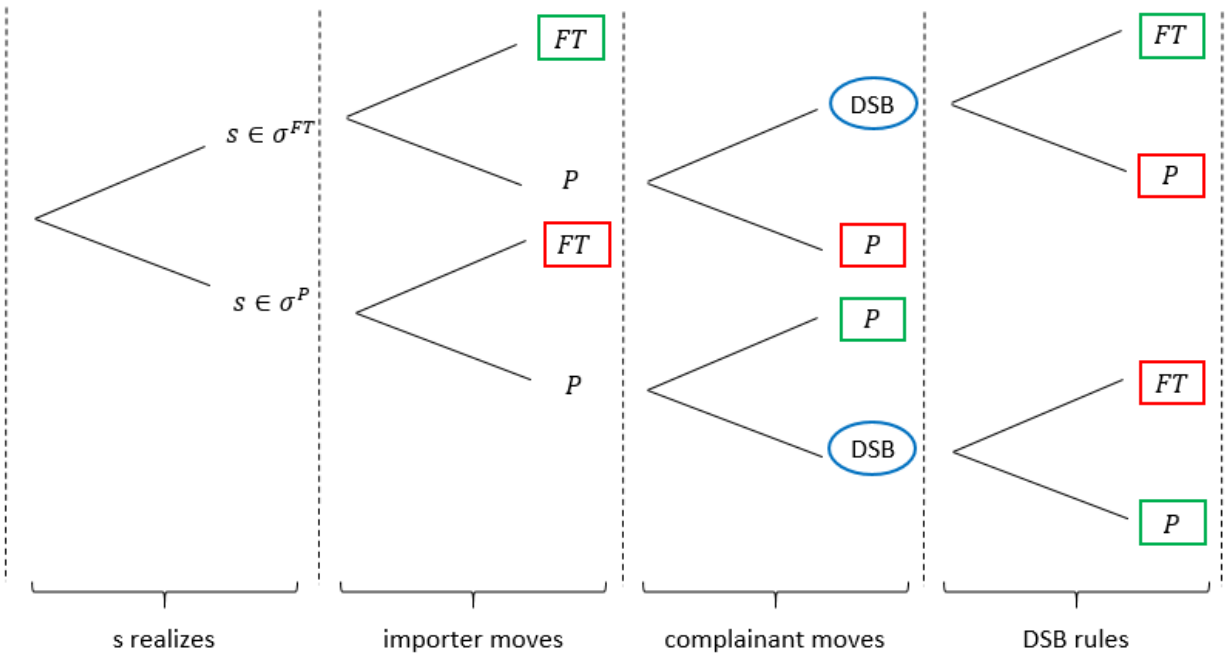

Figure 1: Timing of events (trade agreement)

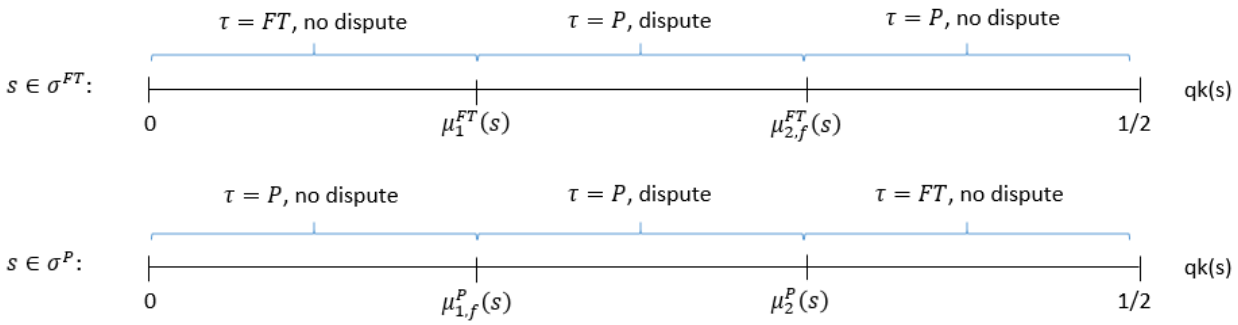

Figure 2a: Illustration of Lemma 1

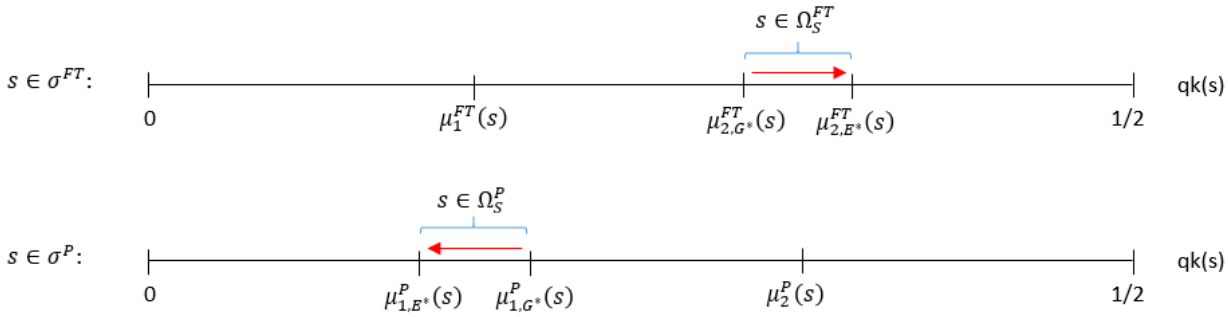

Figure 2b: From SSDS to ESDS

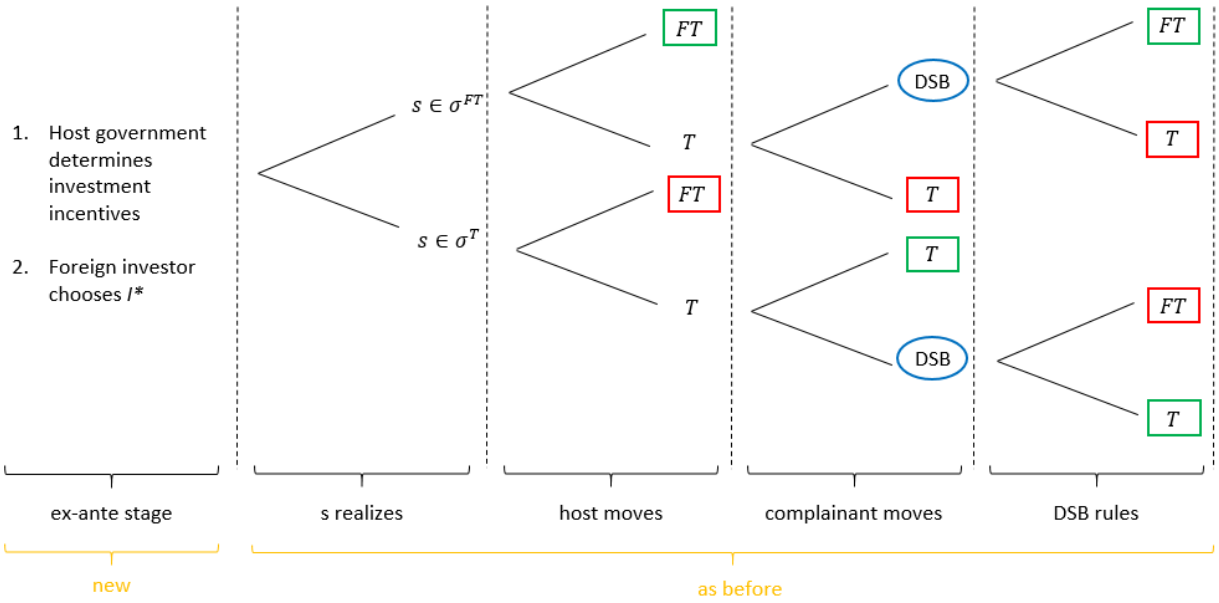

Figure 3: Timing of events (investment agreement)