IFN Working Paper No. 1248, 2018

Investor-State vs. State-State Dispute Settlement

Henrik Horn and Thomas Tangerås

INVESTOR-STATE VS. STATE-STATE DISPUTE SETTLEMENT[1]

Henrik Horn[2]
Research Institute of Industrial Economics (IFN), Stockholm
Bruegel, Brussels
Centre for Economic Policy Research, London

Thomas Tangerås[3]
Research Institute of Industrial Economics (IFN), Stockholm
Energy Policy Research Group (EPRG), University of Cambridge
Program on Energy and Sustainable Development (PESD), Stanford University

February 24, 2022

[2]Address: Research Institute of Industrial Economics (IFN); Box 55665; SE-102 15 Stockholm; Sweden. Email: henrik.horn@ifn.se.
[3]Address: Research Institute of Industrial Economics (IFN); Box 55665; SE-102 15 Stockholm; Sweden. Email: thomas.tangeras@ifn.se.

Abstract

International investment agreements typically permit foreign investors to litigate against host countries (Investor-State Dispute Settlement, ISDS). Yet, common criticism holds that host countries would benefit from allowing only foreign governments to litigate (State-State Dispute Settlement, SSDS). We analyze the negotiated dispute settlement mechanism when SSDS generates political costs that affect the source country's incentives to initiate disputes and the host country's incentives for opportunistic regulation of investments. We show that agreements might rely exclusively on SSDS for given obligations regarding investment protection. However, investment agreements will always include ISDS when countries negotiate both dispute settlement and investment protection.

Contents

1 Introduction

> More importantly, we had situations where real regulation which should be in place, which is bipartisan and in everybody's interest, has not been put in place for fears of ISDS.... (Robert Lighthizer, 2018)[1]

> I oppose the ability of private corporations to attack labor, health, and environmental policies through the Investor-State Dispute Settlement (ISDS) process. (Joseph Biden, 2020)[2]

More than 2 500 *international investment treaties* seek to encourage foreign direct investment by protecting investors against host country policy interventions.[3] Most treaties are bilateral. But it has also become common to combine investment protection with trade liberalization commitments in preferential trade agreements, sometimes involving a large number of member states. The agreements have been criticized in a number of respects in both the policy debate, and in a large academic literature, in law in particular.[4]

The main allegation is that the agreements can induce host country governments to refrain from taking desirable policy measures. Such "regulatory chill" is seen to be caused by a combination of two general features of the agreements. The first is that the substantive obligations in the agreements can be interpreted to request compensation also for what critics consider as legitimate host country policy measures, due to their broad scope. The other feature is that the agreements almost invariably allow private foreign investors to litigate against host country governments, that is, they allow for *Investor-State Dispute Settlement* (ISDS).[5,6] This form of dispute settlement is very rare in international law, since treaties normally allow only signatory states to challenge alleged violations of the treaties by other states—*State-State Dispute Settlement* (SSDS). It is alleged that host countries would be better off without ISDS, since source country governments can be expected to be less prone to challenge legitimate policy measures than private investors.[7]

The investment protection regime is undergoing substantial modifications, both through redraft-

---

[1] Statement made by the Trump administration US Trade Representative before the House Ways and Means Committee on March 21, 2018. https://www.c-span.org/video/?c4719932/brady-lighthizer-isds-discussion.

[2] https://www.uswvoices.org/endorsed-candidates/biden/BidenUSWQuestionnaire.pdf.

[3] Investmentpolicy.unctad.org/international-investment-agreements.

[4] For critical discussion from an academic perspective, see e.g. Stiglitz (2008) and Howse (2017). For a critical policy perspective, see e.g. https://www.isds.bilaterals.org/?lang=en or https://ccsi.columbia.edu/content/investment-law-policy.

[5] The term ISDS is sometimes taken to refer to international investment agreements in general, and sometimes to refer to investor-state arbitration outside domestic legal systems of host countries. This paper refers to ISDS in the literal sense of allowing private investors to bring disputes against host countries. Also, for convenience we use the term "litigate" in the sense of intiating arbitration proceedings, although this is not fully in line with legal terminology.

[6] See UNCTAD (2014) for a comprehensive description of ISDS.

[7] Other observers question the possibility for investors to bypass the domestic legal systems, while still arguing in favor of legal standing of private investors. For instance, more than two hundred academics took this position during the renegotiation of NAFTA; see https://www.citizen.org/system/files/case_documents/isds-law
-economics-professors-letter-oct-2017_2.pdf.

ing of existing agreements, and through novel design of new agreements. One fundamental change is the introduction of *explicit carve-out provisions* that specify circumstances under which host countries can legally intervene without being liable for compensation payments, such as provisions that allow host countries to undertake measures that are necessary to protect human, animal or plant life and health. A second trend is to *reduce the ambit of ISDS*. For instance, the revised version of NAFTA, the United States-Mexico-Canada Agreement (2020), no longer allows for ISDS between Canada and the other two countries, and there is drastically reduced scope for ISDS between the US and Mexico. The investment chapter of the US-Australia Free Trade Agreement (2005) only allows for SSDS, and so does the Japan-UK Comprehensive Economic Partnership Agreement (2020), the post-Brexit EU-UK Trade and Cooperation Agreement (2020), as well as the still negotiated investment chapter in the free trade agreement between Australia and the UK. Another recent change is to let agreements initially only include SSDS while postponing negotiations over whether to include ISDS to a later stage; this has been done in the not yet ratified EU-China Comprehensive Agreement on Investment, and in the very large free trade agreement Regional Comprehensive Economic Partnership (2020).

The intensive policy debate regarding ISDS, and the recent radical changes to the design of international investment agreements, raise a large number of questions regarding the design and performance of investment agreements. The purpose of this paper is to address fundamental question with regard to the choice of dispute settlement mechanisms: How would an exclusion of ISDS affect regulation and investment? Would it benefit the host country by causing less regulatory chill? How would it affect the negotiated investment protection and the implemented protection? If host countries benefit from excluding ISDS, why do negotiated agreements still allow for ISDS to such a large extent? The existing meagre literature on investment agreements cannot yield any insight into these questions. Most papers do no not distinguish between ISDS and SSDS at all. The few papers that do make such a distinction do not allow countries to negotiate the level of investment protection in their agreements (see review below). In contrast, this is the first integrated economic analysis of the scope of their substantive rules and the design of the dispute settlement mechanisms that enforce these provisions in negotiated agreements.

The paper builds on the standard explanation in the legal literature for the inclusion of ISDS in investment agreements, and hence for the difference between the two forms of dispute settlement. As described in Section 2, ISDS *avoids the political and/or diplomatic costs that state-to-state disputes give rise to.* The paper amends this view by the equally natural assumption that costly political tensions also arise if a host country *deliberately violates the agreement* by regulating without paying stipulated compensation, thereby causing the basis for a dispute. It is assumed that both forms of relational costs can depend on the severity of the regulatory shock and the level of protection that the agreement stipulates.

The rest of the paper is structured as follows. Section 3 lays out the economic framework, which in certain respects is a simplified version of the model explored in Horn and Tangerås (2021). In

a first stage, a single investor from the source country undertakes an irreversible investment in a production plant in the host country. Production generates value in Home through job creation, technology transfers, and more, but can also have negative consequences by causing environmental damage or health problems in the host country. We model these negative effects as a stochastic shock, the magnitude of which is revealed only after the investment has been sunk. The shock can be so large that the value of production becomes negative both from a host country and a joint perspective. In the second stage, the host country decides whether to permit production or to shut down the plant through regulatory expropriation.

The interaction between the decisions on investment and regulation generates two distortions. First, the foreign investor may *over- or underinvest* relative to the jointly efficient outcome because the investor disregards all consequences of the investment, both positive and negative, in the host country. Second, the host country tends to *overregulate* since it ignores the consequences for the foreign investor of the decision whether to permit or disallow production. Under plausible assumptions, the countries have mutual interests in improving investment protection because of a first-order effect on the host country through increased investment and a first-order effect on the expected profit of foreign direct investment. A fundamental role of an investment agreement is to implement such protection.

Section 4 introduces an investment agreement, focusing on the standard substantive obligation to compensate investors in case of *regulatory expropriations*—that is, for policy measures with largely the same effect for investors as direct expropriation, but without any formal seizure of assets. The agreement under study specifies *when* regulation is compensable, and *how large* the compensation shall be. Based on the design of actual agreements, we consider a class of *carve-out compensation schemes*: The host country must fully compensate the investor for its foregone operating profit whenever regulation is compensable, but the agreement also establishes a carve-out from the compensation requirement for regulatory shocks that are sufficiently severe. This threshold for the compensation requirement is referred to as the *level of investment protection.* Furthermore, our agreement specifies who has *legal standing*—investors (ISDS) or the source country government (SSDS)—to initiate formal proceedings against the source country regarding violations of the agreement.

It is not obvious how to analyze the claim that host countries, and possibly also source countries, could benefit from excluding ISDS, since the form of dispute settlement is endogenously chosen by the parties when they form an agreement. We will approach the issue from two different angles. Section 5 considers the impact of an exogenous switch from ISDS to SSDS, assuming that the level of investment protection is unaffected. It first derives the equilibrium for an agreement based on ISDS. In this equilibrium, the level of protection that the agreement stipulates is always implemented, since the investor always has an incentive to enforce as much investment protection as possible. The section then performs a parallel analysis for an agreement based on SSDS. The agreement now implements the stipulated investment protection as long as one or more of three incentive

3

compatibility (IC) constraints is fulfilled. The outcome is then precisely the same as under ISDS. However, if all three IC constraints are violated, the host country will illegally regulate without compensation for certain realizations of the regulatory shock. We denote this as *opportunistic regulation* since the host country then exploits that some unlawful regulation will go unchallenged by the source country. In such instances, the host country behavior not only generates political costs in equilibrium, but also causes the expected investment protection to drop below the stipulated level due to opportunistic regulation for certain regulatory shocks, which in turn triggers a reduction in investment.

Drawing on these observations, we show how an exogenous switch to SSDS will induce the host country to regulate opportunistically in certain instances. This *erosion of the stipulated investment protection* can from a host country perspective be viewed as a beneficial reduction of "regulatory chill" from the agreement. But it will come at the cost of lower investment, and it will also expose the host country to expected political costs for opportunistic violations of the agreement. It might still on balance be beneficial for the host country. *Both parties* might prefer SSDS if the host country can compensate the source country for its costs associated with opportunistic regulation. These findings seem at least partly consistent with the criticism against ISDS in the policy debate, but also raise the question: if SSDS is equally good or better than ISDS, why do agreements allow for ISDS to such a high degree?

The analysis has thus far assumed an arbitrary constant level of investment protection. But when forming an agreement, the parties negotiate not only the type of dispute settlement, they also determine the degree of protection that the agreement will give. Section 6 examines how negotiations over both the dispute settlement system and the level of investment protection affect the choice of dispute settlement system and regulatory chill. It first establishes that even if both parties for a given level of investment protection would prefer SSDS to ISDS, *simultaneous negotiation over stipulated investment protection and dispute settlement always yields agreements with ISDS*. Note that we are *not* assuming that the negotiations maximize the parties' joint surplus—if this were the case, an agreement would quite naturally feature ISDS due to the political costs with SSDS. Instead, we consider a general bargaining format that encompasses both the Nash Bargaining Solution and joint welfare optimization as special cases. The negotiation will thus typically not maximize joint surplus due to the distributional conflict between parties. The parties will nevertheless always agree on ISDS. Furthermore, a host country will not use a strong bargaining position to impose SSDS even if it prefers SSDS to ISDS for given investment protection. Instead, it will use its bargaining power to negotiate a level of investment protection that is as close a possible to its preferred level.

Second, to illuminate the recent bargaining practice of agreeing on SSDS while postponing the negotiations regarding ISDS to a later date, Section 6 considers a sequential negotiation of investment protection and dispute settlement. The sequential structure renders it more likely that the agreement will only contain SSDS for an arbitrary level of investment protection, since the host country now is in a position to veto ISDS in the second stage. However, it is shown that if the

4

parties foresee this possibility in the first stage, they will negotiate a level of investment protection such that the host country does not have an incentive to veto ISDS in equilibrium in the second stage negotiation. Hence, *sequential bargaining over investment protection and the form of dispute settlement cannot lead to the exclusion of ISDS in equilibrium.* The sequential negotiation can still lead to a different level of investment protection compared to a simultaneous negotiation. But it is ambiguous whether it will be higher or lower than with ISDS.

Section 7 concludes the paper with a brief summary, and by pointing to some directions for future research. The Appendix provides formal proofs of some of the statements in the main text.

Contribution to the economic literature Markusen (1998, 2001) was among the first to discuss theoretical aspects of investment agreements, and other early contributions were made by e.g. Turrini and Urban (2008), and Aisbett, Karp and McAusland (2010a,b). The recent fierce policy debate has inspired renewed interest in the properties and implications of these agreements. Several theory papers have been published during the last few years, including studies by Konrad (2017), Janeba (2019), Kohler and Stähler (2019), and Horn and Tangerås (2021); Janeba (2020) reviews main themes in this literature. However, none of these papers distinguish between ISDS and SSDS.

The only relevant paper is by Ossa, Staiger and Sykes (2020), who compare differences in dispute settlement procedures between trade and investment agreements. One of those differences relates to the question of who has legal standing to bring a dispute before an arbitration panel. They establish a trade-off between ISDS and SSDS in investment agreements. On the one hand, the increased litigation associated with ISDS reduces inefficient regulatory chill. On the other hand, ISDS increases successful litigation in circumstances where regulation would have been more efficient. The main difference between our paper and Ossa, Staiger and Sykes (2020) is that they treat investment protection as exogenous, whereas we endogenize this aspect of the agreement. Allowing endogenous investment protection has fundamental implications for the desirability of ISDS relative to SSDS.

2 The historical rationale for ISDS

The current investment regime goes back to the early 1960s.[8] Before its creation, foreign investors mainly had to resort to domestic legal systems in host countries to address grievances regarding their treatment. These legal systems were often seen as inadequate in several respects: for instance, national laws did not always reflect the standards of treatment of international investment (or investors) that had been established as customary international law, and host country courts were

---

[8]See e.g. Newcombe and Paradell (2009) and Sykes (2005) for a description of the historical origins of the IIA regime, UNCTAD (2014) for a comprehensive description of the current ISDS regime, and Salacuse (2007), Trevino (2013), Bernasconi-Osterwalder (2014), Roberts (2014), Gertz (2017) and Menon (2018) for discussions of SSDS.

often alleged to be biased against foreign investors in their application of national laws. If these processes failed to give investors the legal treatment they wanted, they had to try to persuade their home governments to exert pressure on the host states. Such "diplomatic protection" often led to political tensions between source and host countries, and in some cases even took the form of literal "gun boat diplomacy."[9] These interventions were costly to both source and host country governments, in a variety of ways. The system also often worked poorly from the point of view of many investors. For instance, investors normally had to exhaust local remedies before requesting support from their governments, which took time. Smaller investors in particular could have problems persuading their governments to intervene. When they did intervene, the governments were in control of the processes, and were also the recipients of any compensation payments.

As a solution to the perceived problems with domestic legal systems in developing countries in particular, developed and developing countries started to form bilateral investment protection treaties in the 1960s. These treaties laid down principles for the required treatment of foreign investments, and also provided for adjudication of disputes through ad hoc constituted arbitration tribunals outside of domestic legal systems. The early treaties only provided for SSDS. The first treaty to include ISDS with unqualified state consent to arbitration came 1969, which thus marks the beginning of the ISDS regime. However, this stipulation became increasingly common in the 1970s, and has been a standard component since the 1990s. For instance, Pohl et al. (2012) found that 93% included ISDS in a sample of 1 660 bilateral investment treaties.

Importantly, several multilateral conventions on international arbitration were formed around the time of the formation of the early treaties. Through the *Convention on the Recognition and Enforcement of Foreign Arbitral Awards* (the "New York Convention") from 1958, signatories undertook to recognize and enforce awards made in other signatory countries; the convention has to date been ratified by 168 countries. Another important step was the formation of the *United Nations Convention on International Trade Law* (UNCITRAL) in 1966 by the General Assembly. Among other things, UNCITRAL developed a set of rules for international arbitration that could be committed to in international investment agreements, and it still performs such a role. A third important step was the *Convention on the Settlement of Investment Disputes between States and Nationals of Other States* (ICSID Convention) in 1966, which sets rules for disputes between its current 155 member states. While UNCITRAL awards are enforced through national courts in accordance with the New York Convention, the ICSID Convention requests contracting states to accept and enforce ICSID awards as if they came from the states' domestic courts. Importantly, due to these multilateral conventions and agreements, *investment agreement have much stronger enforcement mechanisms than almost any other international agreements*, including trade agreements.[10] The ICSID Convention also led to the creation of the *International Centre for Settlement*

---

[9]Kriebaum (2018) gives some drastic examples: the nationalization of the Suez Canal in 1956 led to a military invasion in Egypt by the United Kingdom and France; the Cuba crisis started with nationalizations of US investors by Cuba; and the conflict between the USA and Iran began as a conflict regarding treatment of an oil company.

[10]It is in principle possible for an agreement to allow investors legal standing, while stipulating that disputes have

6

of Investment Disputes (ICSID) in 1966. It is today the most active arbitration institute in terms of the number of investment disputes.

The perceived problems with inadequate and biased host country legal systems could in principle have been addressed through the formation of agreements with SSDS only. The standard explanation by legal scholars and practitioners for the inclusion of ISDS is the desire to "depoliticize" investment disputes. For instance, the longserving Secretary-General of ICSID, Ibrahim Shihata (1986, p. 4) states that the organization:

> ...attempts in particular to 'depoliticize' the settlement of investment disputes.

In their standard text on investment treaties, Dolzer and Schreurer (2012, p. 9) claim:

> From the point of view of member states, one major advantage of the system was that investment disputes would become 'depoliticized' in the sense that they avoided confrontation between home state and host state.

Discussing the extent to which ISDS actually depoliticizes disputes, Paparinskis (2010, p. 271-272) states:

> The contemporary State practice, case law and legal writings consider it almost axiomatic that depoliticization is the purpose of investment protection regime.

Kenneth Vandevelde (1988, p. 258), former US negotiator of investment treaties, writes that ISDS:

> ...ensures investors of a neutral mechanism for settlement of investment disputes that is wholly insulated from the political relationship between the investor's government and the host government.

Ursula Kriebaum (2018, p. 14), member of the Permanent Court of Arbitration, argues that international investment law should be understood

> ....as an instrument for the depoliticization of conflicts that at times had led to fierce confrontations between home and host States of investors.

The legal literature is thus very clear: *depoliticization of disputes was seen as the main advantage of ISDS when it was introduced.* This paper seeks to examine how such a difference between ISDS and SSDS will affect the parties' preferences over the choice of dispute settlement system.

---

to be taken through domestic courts. Symmetrically, it is possible that an agreement only allows contracting states to bring disputes, while at the same time require that they are adjudicated outside domestic legal systems of the host country. We will not distinguish between different legal system for the adjudication of disputes, however.

3   The problems for an investment agreement to address

We first describe the model under the assumption of no investment agreement. The formal setting is in this regard a special case of the framework used in Horn and Tangerås (2021). There are two countries, a host country for foreign direct investment and a source country. There is only one firm in the source country that can undertake such investment, and all investment is undertaken in the host country.[11]

At the outset, the firm makes an irreversible investment $k \geq 0$. The firm's investment cost is a continuous, strictly increasing and weakly convex function $R(k)$ of the size of the investment, and $R(0) = 0$. The investor receives the operating profit $\Pi(k)$ if production is allowed, where $\Pi(k)$ is continuous, strictly increasing and strictly concave in k, and $\Pi(0) = 0$. For the host country, investment creates benefits. The exact nature of these benefits is immaterial, but it could be in terms of consumer surplus, employment, technological spill-overs, learning-by-doing in the work-force, and so forth. After the investment has been undertaken, a shock θ is realized that affects the net benefit to the host country of allowing production. High realizations of θ could represent the arrival of severely adverse information regarding environmental or health consequences of the production process or the goods produced, or other factors affecting the desirability of the investment. The shock is continuously distributed on $[0, \bar{\theta}]$ with cumulative distribution function $F(\theta)$, continuous density $f(\theta) > 0$ and $F(0) = 0$.

Having observed this common-knowledge shock, the host country decides whether to permit or to regulate production. Regulation implies that production is effectively shut down, and thus deprives the firm of its operating profits, causing $\Pi(k) = 0$. This is the sole consequence of regulation for the source country.[12] The host country welfare is $V(k, \theta)$ in case of production, reflecting the net of the positive and negative effects of the investment. The host country welfare function is continuous and differentiable in both arguments. The higher is the realization of θ, the lower is the welfare level of the host country: $V_\theta(k, \theta) < 0$ (subscripts on functional operators denote partial derivatives throughout). $V(k, \theta)$ can be either positive or negative in case of production, and it is zero if there is no production. The marginal net benefit of investment can also be positive or negative, $V_k(k, \theta) \gtrless 0$.[13] To ensure that there is a role to play for investment and regulation, we assume that for every $k > 0$, the host country prefers production if the shock is sufficiently mild, $V(k, 0) > 0$, and prefers regulating if the shock is sufficiently severe, $V(k, \bar{\theta}) < 0$.

The interaction is solved for backwards in standard fashion. The final stage is the host country's decision whether to allow production or to regulate, given investment k and the realized regulatory

---

[11] Including also domestic firms in the industry could create a role for such National Treatment provisions typically contained in investment agreements.

[12] We assume that the source country government abstains from intervening through diplomatic or other means in case the investor is regulated. Thus, there will be no incidence of gunboat diplomacy. This assumption is compatible with a small investor and a political cost that prevents the source country from intervening unless substantial profit is at stake.

[13] The functions $\Pi(k)$, $R(k)$ and $V(k, \theta)$ are assumed to be twice continuously differentiable.

shock θ. Since regulation yields zero welfare for the host country, and $V_\theta < 0$ for all $k > 0$, it is sequentially rational for the host country to allow production whenever $\theta \leq \Theta(k)$, where

$$V(k, \Theta(k)) \equiv 0,$$

and to regulate otherwise. Hence, $F(\Theta(k))$ is the probability that the investment will not be regulated, given that its magnitude is k. This investment "protection" is awarded because it is in the host country's ex post self-interest to allow the firm to produce for mild shocks even without an agreement.

Investment is made prior to the realization of the regulatory shock θ and prior to the regulatory decision. To capture the notion that the investor is small relative to the market, we assume that the investor does not take into consideration how its investment affects the probability of regulation.[14] If the investor expects regulation if and only if $\theta > \theta'$, then its expected profit equals $F(\theta')\Pi(k) - R(k)$, and the optimal investment is given by

$$K(\theta') \equiv \arg\max_{k \geq 0}\{F(\theta')\Pi(k) - R(k)\}.$$

More investment protection obviously translates into more investment, $K_{\theta'}(\theta') > 0$ if $K(\theta') > 0$. Letting tildes denote expected values of variables or functions, the expected net profit is

$$\tilde{\Pi}(\theta') \equiv F(\theta')\Pi(K(\theta')) - R(K(\theta')) \geq 0. \tag{1}$$

Equilibrium investment and investment protection absent an investment agreement is a pair (k_t^0, θ_t^0) given by $k_t^0 = K(\theta_t^0)$ and $\theta_t^0 = \Theta(k_t^0)$, with superscript 0 denoting a non-cooperative outcome under the "null" (no) agreement. Here, θ_t^0 is the sequentially rational investment protection when investment is k_t^0, and the investment k_t^0 is optimal when the investor expects investment protection θ_t^0. Subscript t indicates that this game can have multiple equilibria.[15] Those equilibria can be ranked in increasing order of θ_t^0 (or equivalently k_t^0). For the sake of equilibrium selection we assume that an investor always expects the *maximal* investment protection that is consistent with rational behavior by the host country. We denote the maximal equilibrium absent an investment agreement by (k^0, θ^0), and assume that $\theta^0 < \bar{\theta}$. The following Lemma establishes a fundamental property of regulation when there is no investment agreement in place (see the Appendix for the proof):

---

[14] An alternative approach would be to consider a large number of symmetric investors that all are identically treated by the host and the source country. Each firm would then disregard the impact of its investment on the probability of regulation. Horn and Tangerås (2021) show that the jointly efficient outcome actually can be easier to implement through an investment agreement if firms invest strategically with regard to the host country regulation. But this result requires agreements that are contractually more sophisticated than the type of agreement we will consider here.

[15] Since $\Theta(K(\theta')) \in [0, \bar{\theta}]$ for all θ', it follows that $-\Theta(K(0)) \leq 0$ and $\bar{\theta} - \Theta(K(\bar{\theta})) \geq 0$. By continuity of $\theta' - \Theta(K(\theta'))$ in θ', we can apply the Intermediate Value Theorem to deduce that $\theta_t^0 - \Theta(K(\theta_t^0)) = 0$ for some $\theta_t^0 \in [0, \bar{\theta}]$. Set $k_t^0 = K(\theta_t^0)$. This proves that the game has at least one (pure strategy) equilibrium.

Lemma 1 *If the firm invests $K(\theta') > 0$ in the expectation of investment protection $\theta' \in (\theta^0, \bar{\theta}]$, then $\Theta(K(\theta')) < \theta'$.*

That is, Lemma 1 establishes that all assertions $\theta' > \theta^0$ by the firm about investment protection are too optimistic absent an investment agreement. The equilibrium expected host country welfare equals

$$\tilde{w}^0 \equiv \int_0^{\theta^0} V(k^0, \theta) dF(\theta) \geq 0,$$

Using superscript "*" to denote welfare functions and levels pertaining to the source country, the equilibrium expected source country welfare equals expected investor net profit:

$$\tilde{w}^{*0} \equiv F(\theta^0)\Pi(k^0) - R(k^0) \geq 0.$$

There are two distortions at work, both of which are fundamental to the problems that investment agreements seek to remedy. First, the investor disregards both the positive and the negative externalities from the investment that occur in the host country. Second, the host country disregards foreign investor profit when deciding whether to regulate. Each distortion is straightforward on its own, but their interaction is more involved. Still, unilateral investment decisions by foreign investors and unilateral regulatory decisions by host countries typically entail distortions of both investment and regulation. In particular, we assume that the host country would unilaterally benefit from being able to increase investment protection above θ^0 by an associated increase in investment:

$$\int_0^{\theta'} V(K(\theta'), \theta) dF(\theta) > \tilde{w}^0 \text{ for some } \theta' \in (\theta^0, \bar{\theta}]. \tag{2}$$

Sufficient conditions for this property to hold are that $k^0 > 0$, the marginal host country benefit of investment is weakly decreasing in the regulatory shock, $V_{k\theta}(k, \theta) \leq 0$, and that the marginal welfare of investment is strictly positive evaluated at (k^0, θ^0): $V_k(k^0, \theta^0) > 0$. Since the source country is only concerned with investor profits in this setting, and the marginal net profit with respect to the level of protection is positive,

$$\tilde{\Pi}_\theta(\theta') = f(\theta')\Pi(K(\theta')) > 0, \ K(\theta') > 0 \tag{3}$$

the two countries have a mutual interest in increasing investment protection above θ^0. A basic aim of an investment agreement is to achieve this improvement in investment protection.

4 An investment agreement

We will now introduce our formalization of an investment agreement, and the difference between ISDS and SSDS.

4.1 Formalizing core features of an agreement

As mentioned above, there are considerable similarities across actual investment agreements.[16] There is no direct contracting on investment levels nor on regulation, nor do the agreements include any commitments regarding direct subsidies to, or taxes on, investors. But virtually all agreements contain provisions that require compensation in case of both *direct* and *regulatory (indirect) expropriation*. Direct expropriation occurs when a host country seizes a foreign investor's assets. Regulatory expropriation is a policy intervention by the host country with largely the same effect on investors as direct expropriation, but without any formal taking of assets. The ambit of regulatory expropriation clauses has been a frequent source of contention in both the case law and in the policy debate.

The scope of a regulatory expropriation clause can be limited in at least two respects. The first concerns the interpretation of the notion of regulatory expropriation. Some panels have taken the view that all that matters is the effect of the government measures on investors; the intent behind the measure is irrelevant. But more recent panels have often interpreted the agreements to impose less severe restrictions on host countries, emphasizing the "police powers exemption" in international law that allows states to protect public welfare.[17] The other, and increasingly common, limitation is through *explicit carve-outs* from compensation requirements. For instance, a treaty might state that its stipulations shall not preclude the parties from adopting or enforcing any measure that is "necessary to protect human, animal or plant life or health", provided that the measure does not constitute "disguised protection", or "arbitrary or unjustifiable discrimination."

To capture these features of actual treaties, we formalize an agreement as consisting of two components. The first component is a specification of the amount of compensation the investor will receive in case of regulation, as a function of underlying circumstances. The compensation requirements in most actual investment agreements share a number of features that are central from a contractual point of view:

1. All compensation payments are from the host country to foreign investors and are only required subsequent to regulatory intervention.

2. There are no payments to or from outside parties.

3. There can be carve-outs from compensation requirements for certain types of regulatory measures.

4. Any compensation must equal foregone operating profits.[18]

---

[16] See e.g. Dolzer and Schreuer (2012) for an introduction to International Investment Law.

[17] Often cited examples of the former approach are the panels in *Metalclad v. Mexico* (1997), and *TECMED v. Mexico* (2003). Examples of the latter approach, which has also been adopted by a number of later panels, are *Methanex v. United States* (2005), and *Philip Morris v. Uruguay* (2017).

[18] Some panels have used different criteria, such as the magnitude of the investors' investment.

We thus assume that an agreement specifies the following *carve-out compensation scheme:*

$$T(k,\theta,\hat{\theta}) \equiv \begin{cases} \Pi(k) & \text{if } \theta \leq \hat{\theta} \\ 0 & \text{if } \theta > \hat{\theta} \end{cases} \tag{4}$$

where $T(k,\theta,\hat{\theta})$ is the amount that the host country shall pay the investor in case of regulation. Compensation is hence required if the regulatory shock θ is weaker than a threshold value $\hat{\theta}$, but not for shocks that are more severe than $\hat{\theta}$. Furthermore, whenever regulation is compensable, this compensation should equal the firm's foregone operating profit. We will refer to $\hat{\theta}$ as the *stipulated level of protection* of the agreement, since full compensation is requested for all realizations of θ smaller or equal to $\hat{\theta}$, whereas the investor is not entitled to any compensation for regulation that occurs if $\theta > \hat{\theta}$.

The second standard component of the agreement is a specification of *who has legal standing to initiate a dispute.* As noted above, most agreements allow both investors and states to litigate. The agreement we will consider thus also specifies whether the source country government only (SSDS), or also the investor (ISDS), can pursue disputes.

4.2 The sequence of events with an agreement

To capture the very long-term nature of investment agreements, we consider agreements that are formed at the outset of the interaction. The events then unfold as follows:

1. The firm invests k.

2. A regulatory shock θ is realized.

3. The host country decides whether to:
   - allow production;
   - regulate with immediate compensation; or
   - regulate without any associated compensation.

4. If the host country regulates without respecting a compensation requirement, the investor (ISDS) or the source country government (SSDS), decides whether:
   - to let the violation go unchallenged; or
   - to enforce the compensation payment through formal arbitration.

We assume that a dispute court can perfectly observe whether uncompensated regulatory intervention by the host country amounted to illegal indirect expropriation ($\theta \leq \hat{\theta}$) or a legitimate policy intervention under the agreement ($\theta > \hat{\theta}$). A common knowledge shock obviously simplifies the analysis, but one could introduce adjudication with noisy signals. Such an extension would add

12

substantial complexity to the analysis, but with uncertain gains. However, see the discussion in Section 7.

A second simplifying assumption is that there are no legal process costs from pursuing disputes. There are of course in actuality often substantial costs involved. However, there is no obvious reason why they would differ significantly between private investors and governments, and governments often rely on counsel from law firms. Also, the inclusion of such costs would not formally affect results in our complete information setting under the additional assumption that the losing party is liable for all litigation costs.

A central issue in what follows will be the extent to which an agreement *implements* the stipulated level of protection $\hat{\theta}$. Investor beliefs about investment protection determines how much they invest. The magnitude of investment and the realization of the regulatory shock affect the host country's incentive to regulate and the source country's incentive to enforce the stipulated protection under SSDS. Since we are interested in sequentially rational investments, expectations concerning investment protection must be fulfilled in equilibrium. In case the agreement stipulates a non-implementable level of protection, the investor always expects the maximal investment protection that is consistent with rational behavior by the host country, as per our assumption regarding equilibrium selection.

4.3 Formalizing the political costs from violations and enforcement of an agreement

It is not self-evident how to bring the benefits of depoliticization into an economic analysis. Governments typically interact in a wide range of areas. Some are of an economic nature, but countries also cooperate on national security, money laundering, drug enforcement, sharing of air space and innumerous other issues. Political costs can thus take a multitude of forms, many of which fall outside the realm of standard economic analysis. What they have in common however, is that conflicts between governments in one area can sour the relationship in other areas, to the detriment of both parties. We do not see any reason to focus on a specific relational friction from investment disputes, so we will represent these "political" costs in general reduced forms. The main assumption is that they do not affect the functions $\Pi(k)$ and $V(k, \theta)$, but enter quasi-linearly.

We will assume there can be two types of relational costs from an investment conflict. In line with the notion in the legal literature that was described in Section 2, we assume that litigation by the *source country government* regarding uncompensated regulation by the host country when $\theta \leq \hat{\theta}$ (unlawful regulation), will expose the source country to the political cost $N^*(\theta, \hat{\theta})$, and the host country to the political cost $N(\theta, \hat{\theta})$. These costs do not arise when the *investor* enforces the agreement. For the source country, this "enforcement cost" is lower, the lower is θ, since the source country is then challenging a more blatantly opportunistic behavior by the host country; specifically, $N^*(\theta, \hat{\theta})$ is strictly positive, continuous in both its arguments, and strictly increasing in θ for all

$\theta < \hat{\theta}$. As it will turn out, the properties of N are immaterial for the outcome of the interaction.

The other form of relational cost arises when a host country regulates without paying required compensation. These costs are not (to the best of our knowledge) discussed in the legal literature on the ISDS/SSDS distinction, but they seem highly natural to include in an economic analysis that focuses on the relational costs of an investment agreement. We will thus let $M(\theta, \hat{\theta})$ represent the cost for the host country for such aggressive behavior. An illegal regulation is plausibly seen as more confrontational, the weaker is the regulatory shock that the host country is exposed to. Hence, we assume that $M(\theta, \hat{\theta})$ is strictly decreasing in θ, and continuous in all arguments, for $\theta \le \hat{\theta}$. The cost is the same under ISDS and SSDS and arises regardless of whether the host country actions actually triggers litigation. But no such costs arise if the host country either permits production, regulates with immediate compensation payment $\Pi(k)$ for shocks $\theta \le \hat{\theta}$, or regulates without compensation for shocks $\theta > \hat{\theta}$, since in all three instances the host country abides by the terms of the agreement. Since regulation without required compensation sours the relationship between the governments, it will expose also the source country to a political cost $M^*(\theta, \hat{\theta})$ with qualitatively the same properties as $M(\theta, \hat{\theta})$. The political cost $M^*(\theta, \hat{\theta})$ will not affect the incentives facing either country, but will still affect source country welfare.[19]

5 Excluding ISDS for given investment protection

In what follows we will examine the incentives for the source and the host country with regard to the choice of dispute settlement system, and the implications for regulatory chill. It is not clear how to examine the notion that ISDS is harmful to host countries, since the form of dispute settlement is endogenously chosen by the parties, and thus cannot readily be switched on and off. But to shed some light on this popular notion, we will start by examining the consequences of *exogenously* excluding ISDS from an agreement that includes SSDS for a *given arbitrary* agreement $\hat{\theta}$. This will serve to identify some basic aspects of how the form of dispute settlement system affects the parties. The analysis might also capture situations where an agreement no longer reflect the preferences of the parties since the underlying conditions have changed. For instance, the host country might have developed economically, and thus be in less need of foreign investment, or there might have been political changes in the host country. Another possibility would be that countries initially over-estimated the effect of the agreement on investment, perhaps by failing to realize the effects of globalization on competition for investment across countries. All these empirically plausible scenarios could cause an agreement to become poorly adopted to current preferences of the parties.

The analysis will be performed in three steps: Sections 5.1 and Section 5.2 derive the equilibrium outcome for the negotiated level of investment protection, investment and regulation, with ISDS and SSDS, respectively. Drawing on these sections, Section 5.4 establishes some fundamental differences between the functioning of the two dispute settlement systems.

---

[19] We could also assume that the political costs are functions of k. This would not affect the analaysis below, however.

5.1 ISDS

Assume that the countries have an agreement that allows for ISDS. The agreement stipulates the compensation function $T(k, \theta, \hat{\theta})$ defined in (4), where $\hat{\theta} \in (\theta^0, \bar{\theta}]$ is some exogenous arbitrary level of investment protection. Observe that $\hat{\theta} > \Theta(K(\hat{\theta}))$ by Lemma 1. Under incomplete investment protection, $\hat{\theta} < \bar{\theta}$, the agreement allows the host country to regulate without compensation for sufficiently severe shocks, $\theta \in (\hat{\theta}, \bar{\theta}]$, and the host country will prefer to do so if the firm has invested $K(\hat{\theta})$. If the host country regulates for $\theta \leq \hat{\theta}$ it is obliged to pay compensation under the terms of the agreement.

Suppose the host country violates the agreement by regulating without compensation for $\theta \leq \hat{\theta}$. With ISDS, the investor will always enforce any stipulated payments by the assumption that doing so is without cost. Consequently, the host country will be forced to pay compensation in the end. It then ends up financially in the same position as with immediate compensation payment, except it has now incurred the additional political cost $M(\theta, \hat{\theta})$ by violating the agreement. Therefore, the host country will either allow production or regulate with immediate compensation for all shocks $\theta \leq \hat{\theta}$ under ISDS.

We assume that the host and source country welfare functions are quasi-linear in a transfer payment $T(k, \hat{\theta}, \theta)$. The host country then prefers regulation if and only if is better to pay $\Pi(k)$ in compensation to avoid welfare $V(k, \theta) < 0$ associated with allowing production, that is, if and only if $V(k, \theta) < -\Pi(k)$. There is thus a threshold for the regulatory shock $\Theta^J(k)$ defined by

$$V(k, \Theta^J(k)) \equiv -\Pi(k)$$

if $V(k, \bar{\theta}) \leq -\Pi(k)$ and by $\Theta^J(k) = \bar{\theta}$ otherwise, such that the host country is willing to regulate even if it has to pay compensation for $\theta > \Theta^J(k)$.

The investor behaves non-strategically vis-à-vis host country regulatory decisions. This might seem as an unappealing assumption with a single investor. But the assumption regarding a single investor is not meant to capture a concentrated industry. It is maintained to avoid e.g. the coordination problems that would arise when more than one investor is affected by the same measure. While such aspects can be relevant in certain instances, they do not seem to be of first-hand importance for basic questions regarding the difference between ISDS and SSDS. The investor will therefore invest $K(\hat{\theta})$ if the expectation is either to gain permission to produce or to be regulated with full compensation for all $\theta \leq \hat{\theta}$, and to be regulated without compensation for $\theta > \hat{\theta}$. The following result is immediate:

Lemma 2 *Any investment agreement that stipulates investment protection $\hat{\theta} \in (\theta^0, \bar{\theta}]$, compensation $T(k, \theta, \hat{\theta})$ and ISDS, has the following properties:*
(i)   The firm invests $K(\hat{\theta})$.
(ii)   The host country allows production for all $\theta \leq \min\{\hat{\theta}; \Theta^J(K(\hat{\theta}))\}$.

(iii) The host country lawfully regulates by paying immediate compensation if investment protection is extensive, $\hat{\theta} > \Theta^J(K(\hat{\theta}))$, and $\Theta^J(K(\hat{\theta})) < \theta \leq \hat{\theta}$.

(iv) The host country lawfully regulates without paying compensation if the agreement features incomplete investment protection and $\hat{\theta} < \theta \leq \bar{\theta}$.

The host country expected welfare in an agreement with the stipulated protection level $\hat{\theta}$ and ISDS, $\tilde{W}(\hat{\theta}, I)$, is given by

$$\tilde{W}(\hat{\theta}, I) \equiv \begin{cases} \int_{\underline{\theta}}^{\hat{\theta}} V(K(\hat{\theta}), \theta) dF(\theta) & \text{if } \hat{\theta} \leq \Theta^J(K(\hat{\theta})) \\ \int_{\underline{\theta}}^{\Theta^J(K(\hat{\theta}))} V(K(\hat{\theta}), \theta) dF(\theta) - [F(\hat{\theta}) - F(\Theta^J(K(\hat{\theta})))]\Pi(K(\hat{\theta})) & \text{if } \hat{\theta} > \Theta^J(K(\hat{\theta})). \end{cases}$$
(5)

Host country expected welfare depends on whether compensation payments can arise in equilibrium. If stipulated investment protection is limited in the sense that $\hat{\theta} \leq \Theta^J(K(\hat{\theta}))$, the host country will allow production and obtain welfare $V(K(\hat{\theta}), \theta)$ for mild shocks $\theta \leq \hat{\theta}$, and it will lawfully regulate without compensation payments and obtain zero welfare for $\theta > \hat{\theta}$. These properties yield the expression on the first row of (5). The second row of (5) characterizes the expected host country welfare under extensive investment protection, $\hat{\theta} > \Theta^J(K(\hat{\theta}))$. The host country will permit production and obtain welfare $V(K(\hat{\theta}), \theta)$ for $\theta \leq \Theta^J(K(\hat{\theta}))$; it will regulate and pay full compensation for all shocks $\Theta^J(K(\hat{\theta})) < \theta \leq \hat{\theta}$, in which case its welfare will be $-\Pi(K(\hat{\theta}))$; and it will regulate without compensation payments for all shocks $\theta > \hat{\theta}$, in which case host country welfare will be zero. The term in square brackets on the second row of (5) is the probability that the host country will pay compensation under the agreement.

The source country expected welfare with ISDS, $\tilde{W}^*(\hat{\theta}, I)$, equals the expected investor net profit with ISDS, $\tilde{\Pi}(\hat{\theta})$, since the host country never causes any political costs by breaching the agreement. The firm is either allowed to produce or is regulated with full compensation for $\theta \leq \hat{\theta}$, and regulated without compensation for $\theta > \hat{\theta}$. The source country expected welfare in an agreement with ISDS therefore equals

$$\tilde{W}^*(\hat{\theta}, I) \equiv \tilde{\Pi}(\hat{\theta})$$

regardless of whether there are equilibrium compensation payments. There is scope for such an agreement by $\tilde{W}(\hat{\theta}, I) > \tilde{w}^0$ and $\tilde{W}^*(\hat{\theta}, I) > \tilde{w}^{*0}$ for some $\hat{\theta} > \theta^0$; see (2) and (3).

It is straightforward to see why there might be political opposition in a host country to an agreement that features investment protection $\hat{\theta} > \theta^0$ sustained by ISDS. Assume that the firm has invested $K(\hat{\theta})$. If a shock θ in the range $(\Theta(K(\hat{\theta})), \min\{\hat{\theta}; \Theta^J(K(\hat{\theta}))\}]$ is realized, the host country will find it ex post optimal to allow production instead of regulating, to avoid having to pay compensation to the foreign investor. The agreement will thus cause *domestic regulatory chill* that can be seen to be driven by the credible threat of litigation by the investor that ISDS enables. Second, if there is a regulatory shock θ in the range $(\Theta^J(K(\hat{\theta})), \hat{\theta}]$, the host country will find it ex post optimal to regulate despite having to pay compensation, since it would otherwise face ISDS

litigation. Such a payment to a foreign investor for the right to avoid the consequences a rather severe shock could easily be seen as unreasonable. Third, in these situations the basic purpose of the agreement seems to have failed by generating excessive investment from the viewpoint of the host country. Experiences from these types of situations might plausibly drive the demands for the removal of ISDS from investment agreement.

5.2 SSDS

Consider again an agreement featuring the stipulated compensation $T(k, \theta, \hat{\theta})$, but now with SSDS. From an investor's point of view it is still immaterial whether its revenue is received as operating profit or as compensation for regulation. However, what matters now is not the stipulated, but the *implemented* level of protection. Under SSDS, the two are not always the same.

5.2.1 Enforcement and regulation incentives

Assume that the firm has invested $K(\hat{\theta})$ in the belief that the agreement will implement the stipulated level of protection $\hat{\theta}$. The investor then expects to receive either operating profit $\Pi(K(\hat{\theta}))$ or the same amount as compensation in case of regulation for all $\theta \leq \hat{\theta}$, but no compensation for regulation for more severe shocks $\theta > \hat{\theta}$. Enforcement of the agreement through SSDS only matters for shocks $\theta \in (\Theta(K(\hat{\theta})), \hat{\theta}] \equiv \Lambda(\hat{\theta})$ because the host country will always allow production for shocks $\theta \leq \Theta(K(\hat{\theta}))$ even absent any agreement. The investor will recover its operating profit for all shocks $\theta \in \Lambda(\hat{\theta})$ if the compensation payment is sufficiently large relative to the source country political enforcement cost, that the source country will always use SSDS to challenge any decision by the host country to regulate, without respecting a stipulated compensation payment. Formally, this requires

$$\Pi(K(\hat{\theta})) \geq N^*(\hat{\theta}, \hat{\theta}). \tag{6}$$

The host country then rationally foresees that the source country will enforce the agreement for any uncompensated regulation that occurs for shocks $\theta \leq \hat{\theta}$. Consequently, the host country honors the agreement to avoid the unnecessary political cost $M(\theta, \hat{\theta})$, just as in the case of ISDS.

Things are more complicated if the source country political enforcement cost is so large that the investor cannot always count on investment protection through SSDS: $\Pi(K(\hat{\theta})) < N^*(\hat{\theta}, \hat{\theta})$. This situation leaves the host country with three possibilities:

- Allow production and achieve welfare $V(K(\hat{\theta}), \theta)$.

- Regulate with immediate compensation payments and get $-\Pi(K(\hat{\theta}))$.

- Regulate without paying compensation and get $-M(\theta, \hat{\theta})$; which we denote *opportunistic regulation* since the regulation exploits the lack of enforcement by the source country.

Notwithstanding the lack of source country enforcement, the agreement will implement $\hat{\theta}$ if the host country political cost of violating the agreement is sufficiently high. There are two channels through which the host country political costs affect investment protection. One possibility is that the host country *prefers immediate compensation to opportunistic regulation.* This condition is satisfied for all $\theta \in \Lambda(\hat{\theta})$ if

$$M(\hat{\theta}, \hat{\theta}) \geq \Pi(K(\hat{\theta})). \tag{7}$$

This condition states that the political cost of violating the agreement is sufficiently large relative to the cost of immediate compensation even for the shock $\theta = \hat{\theta}$, that immediate compensation is preferable to opportunistic regulation.

The second possibility is that the host country *prefers production to opportunistic regulation.* This condition is satisfied for all $\theta \in \Lambda(\hat{\theta})$ if

$$V(K(\hat{\theta}), \hat{\theta}) \geq -M(\hat{\theta}, \hat{\theta}). \tag{8}$$

Given these three incentive constraints, we can generally distinguish between two qualitatively different sets of circumstances under SSDS.

5.2.2 Stipulated protection is implemented

It suffices that any of the three incentive compatibility (IC) constraints (6)-(8) is fulfilled for a stipulated investment protection level $\hat{\theta}$ to be implemented under SSDS, if the firm has invested $K(\hat{\theta})$. We let

$$A \equiv \{\hat{\theta} \in (\theta^0, \bar{\theta}]: \text{at least one of (6), (7) and (8) holds}\}$$

be the set of stipulated investment protection levels that can be implemented under SSDS. Potentially, $A = (\theta^0, \bar{\theta}]$, in which case there is no real difference between SSDS and ISDS.

One possibility is that the political costs associated with SSDS do not affect implementation of the agreement:

Lemma 3 *Consider an investment agreement that stipulates investment protection $\hat{\theta} \in (\theta^0, \bar{\theta}]$, compensation $T(k, \theta, \hat{\theta})$ and SSDS. Any such agreement where $\hat{\theta} \in A$, has the same properties as an agreement with ISDS, as characterized in Lemma 2.*

Hence, for $\hat{\theta} \in A$, either the source country enforces the agreement, the host country allows production for $\theta \leq \hat{\theta}$, or the host country regulates with immediate compensation for $\theta \leq \hat{\theta}$.

As noted above, a standard notion holds that enforcement with SSDS gives rise to political costs. But as we have argued, political costs for violations of an agreement seem equally plausible as political costs of enforcing agreements in case of violations. The existence of host country political costs can have real impact on the outcome with SSDS:

18

Observation 1 *With SSDS, host country political costs from violations of the agreement ($M >$ 0) might prevent source country political enforcement costs ($N^* > 0$) from undermining the agreement.*

If an agreement with SSDS implements the stipulated level of protection $\hat{\theta}$, the host country either allows production or regulates with immediate compensation for all $\theta \leq \hat{\theta}$, and it regulates without compensation for all $\theta > \hat{\theta}$. The implementation of the agreement does not impose any political costs on either country, since the host country honors the stipulations in the agreement, and the source country never has to challenge the host country decisions. The expected welfare in both countries then is exactly the same as under ISDS, $\tilde{W}(\hat{\theta}, S) = \tilde{W}(\hat{\theta}, I)$ and $\tilde{W}^*(\hat{\theta}, S) = \tilde{W}^*(\hat{\theta}, I)$ for all $\hat{\theta} \in A$.

5.2.3 Stipulated protection cannot be implemented

The second possibility is that $\hat{\theta} \in A^c \equiv (\theta^0, \bar{\theta}] - A \neq \emptyset$, in which case *all* three IC constraints (6)-(8) are violated for the stipulated level of investment protection. In this case, $\hat{\theta}$ cannot be implemented subsequent to investment $K(\hat{\theta})$ by the firm, implying that actual investment protection will differ from what is stipulated. The rational investor will then correctly anticipate less investment protection than the stipulated level $\hat{\theta}$, and invest accordingly. Assume that the firm invests $K(\theta')$ in the subjective belief that it will receive investment protection $\theta' \leq \hat{\theta}$.

Let $L(\theta', \hat{\theta})$ be the actual level of investment protection implemented by the host country if the firm invests $K(\theta')$ and the agreement stipulates investment protection $\hat{\theta}$. The expected net profit then equals $F(L(\theta', \hat{\theta}))\Pi(K(\theta')) - R(K(\theta'))$. Hence, beliefs and actions form a rational expectations equilibrium only if $\theta' = L(\theta', \hat{\theta})$. If the firm optimistically believes in the stipulated level $\hat{\theta}$ and invests $K(\hat{\theta})$ accordingly, then opportunistic regulation reduces actual investment protection below the stipulated level for $\hat{\theta} \in A^c$: $L(\hat{\theta}, \hat{\theta}) < \hat{\theta}$. The other polar case is of a pessimistic firm that believes that any agreement with stipulated protection $\hat{\theta} \in A^c$ will offer no more investment protection than the situation without any agreement, $\theta' = \theta^0$. The firm consequently invests $K(\theta^0) = k^0$. In this case,

$$V(k^0, \theta^0) = 0 > -M(\theta^0, \hat{\theta})$$

implies that the firm will receive strictly more investment protection $L(\theta^0, \hat{\theta})$ than θ^0, because the political cost $M(\theta, \hat{\theta})$ of opportunistic regulation has a disciplining effect on the host country for a range of mild shocks $\theta > \theta^0$. The pessimistic belief thus understates actual investment protection under SSDS. As $\theta^0 - L(\theta^0, \hat{\theta}) < 0$ and $\hat{\theta} - L(\hat{\theta}, \hat{\theta}) > 0$, $\theta' = L(\theta', \hat{\theta})$ for some $\theta' \in (\theta^0, \hat{\theta})$ by the Intermediate Value Theorem. There can be multiple such fixed points, so let $G(\hat{\theta}) \in (\theta^0, \hat{\theta})$ be the maximal level of protection in the set of fixed points. We verify the following statement in the Appendix:

Lemma 4 *Consider an investment agreement that stipulates investment protection* $\hat{\theta} \in (\theta^0, \bar{\theta}]$, *compensation* $T(k, \theta, \hat{\theta})$ *and SSDS. Any such agreement with* $\hat{\theta} \in A^c$ *has the following properties:*

(i)   The firm invests $K(G(\hat{\theta}))$.

(ii)   The host country allows production if $\theta \leq \min\{G(\hat{\theta}); \Theta^J(K(G(\hat{\theta})))\}$.

(iii) The host country regulates with immediate compensation payment if $\Theta^J(K(G(\hat{\theta}))) < \theta \leq G(\hat{\theta})$.

(iv)   The host country regulates opportunistically if $G(\hat{\theta}) < \theta \leq \hat{\theta}$.

(v)   The host country lawfully regulates without compensation payment if the agreement stipulates incomplete investment protection, and $\hat{\theta} < \theta \leq \bar{\theta}$.

If an agreement with SSDS cannot implement $(K(\hat{\theta}), \hat{\theta})$, that is, if $\hat{\theta} \in A^c$, the host country's expected welfare equals

$$\tilde{W}(\hat{\theta}, S) \equiv \tilde{W}(G(\hat{\theta}), I) - \int_{G(\hat{\theta})}^{\hat{\theta}} M(\theta, \hat{\theta}) dF(\theta). \tag{9}$$

The source country's corresponding expected welfare is

$$\begin{aligned} \tilde{W}^*(\hat{\theta}, S) &\equiv \tilde{W}^*(G(\hat{\theta}), I) - \int_{G(\hat{\theta})}^{\hat{\theta}} M^*(\theta, \hat{\theta}) dF(\theta) \\ &= \tilde{\Pi}(G(\hat{\theta})) - \int_{G(\hat{\theta})}^{\hat{\theta}} M^*(\theta, \hat{\theta}) dF(\theta) \end{aligned} \tag{10}$$

An agreement based on SSDS causes several problems if the political cost of violating the terms of the agreement and the enforcement costs are such that SSDS cannot implement the stipulated level of investment protection. SSDS reduces investment protection $G(\hat{\theta})$ below the stipulated level $\hat{\theta}$, and thereby reduces investment to $K(G(\hat{\theta})) < K(\hat{\theta})$. Moreover, this equilibrium is sustained by opportunistic regulation for shocks $\theta \in (G(\hat{\theta}), \hat{\theta}]$, the expected political cost of which is captured by the second term in each of the above expressions.

5.3   Bargaining format

Negotiations over the agreement occur at the outset of the interaction, before the investment stage. We will not employ any specific bargaining solution, but instead assume that the bargaining outcome maximizes the continuous bargaining function $B(\hat{\theta}, \delta) \equiv \hat{B}(\tilde{W}(\hat{\theta}, \delta), \tilde{W}^*(\hat{\theta}, \delta))$, with $\delta = I$ indicating that the agreement allows for ISDS and $\delta = S$ that there is only SSDS. The function $\hat{B}$ is taken to be strictly increasing in each of its arguments and differentiable everywhere, $\hat{B}_{\tilde{W}} > 0$, $\hat{B}_{\tilde{W}^*} > 0$. A special case of this bargaining problem is that the outcome maximizes joint welfare. However, we are mainly interested in situations where distributional conflicts between the countries can affect the design of the agreement.[20] The negotiated outcome is also assumed to fulfil countries' participation

---

[20]We could additionally define $\hat{B}$ over the expected welfare levels absent an agreement, $\tilde{w}^0$ and $\tilde{w}^{*0}$, in which case another special case could be the Nash Bargaining outcome.

constraints,

$$\tilde{W}(\hat{\theta}, \delta) \geq \tilde{w}^0 \text{ and } \tilde{W}^*(\hat{\theta}, \delta) \geq \tilde{w}^{*0}. \tag{11}$$

and at least one country must strictly benefit for an agreement to be formed.

5.4 Consequences of excluding ISDS

Consider an agreement that stipulates some level of investment protection $\hat{\theta} \in (\theta^0, \bar{\theta}]$ and allows both ISDS and SSDS. Assume that the agreement is redrafted to only allow for SSDS, while leaving investment protection unchanged at $\hat{\theta}$. If $\hat{\theta} \in A$, then sufficient IC constraints for implementation of $\hat{\theta}$ are fulfilled also under SSDS. This implementation can rely on a credible threat of source country enforcement or on substantial political costs of opportunistic regulation for the host country. The outcome then remains fundamentally unaffected despite the fact the exclusion of ISDS forces the parties to rely on a dispute settlement mechanism that yields political costs if activated. Hence, there is nothing to gain for any party from removing ISDS from the agreement if $\hat{\theta} \in A$.

The more interesting cases arise when $\hat{\theta} \in A^c$, so that an agreement with SSDS cannot implement $\hat{\theta}$. Removing ISDS from the agreement will then reduce the implemented level of investment protection from $\hat{\theta}$ to $G(\hat{\theta})$ and investment from $K(\hat{\theta})$ to $K(G(\hat{\theta}))$ as a consequence of the change in investment protection. Enforcement of $G(\hat{\theta})$ will in some instances lead to opportunistic regulation by the host country, which creates political costs in both countries. For the source country, the effect of removing ISDS is strictly negative:

$$\tilde{W}^*(\hat{\theta}, S) - \tilde{W}^*(\hat{\theta}, I) = \tilde{\Pi}(G(\hat{\theta})) - \tilde{\Pi}(\hat{\theta}) - \int_{G(\hat{\theta})}^{\hat{\theta}} M^*(\theta, \hat{\theta}) dF(\theta) < 0. \tag{12}$$

Not only does removal of ISDS reduce expected net profit, as represented by the first two terms above, it also leads to political costs because of opportunistic regulation, which is the third term. The consequences for the host country are ambiguous, however:

$$\tilde{W}(\hat{\theta}, S) - \tilde{W}(\hat{\theta}, I) = \tilde{W}(G(\hat{\theta}), I) - \tilde{W}(\hat{\theta}, I) - \int_{G(\hat{\theta})}^{\hat{\theta}} M(\theta, \hat{\theta}) dF(\theta) \gtrless 0. \tag{13}$$

Removing ISDS from the agreement induces the host country to regulate opportunistically in certain instances, and this will lead to political costs, as given by the third term above. However, the reduction in implemented investment protection from $\hat{\theta}$ to $G(\hat{\theta})$ can have a negative or positive effect on the host country expected welfare from production. Let θ^U be the level of investment protection that maximizes host country expected welfare with ISDS:

$$\theta^U \equiv \arg \max_{\hat{\theta} \in (\theta^0, \bar{\theta}]} \tilde{W}(\hat{\theta}, I). \tag{14}$$

21

If the exogenously given $\hat{\theta}$ is larger than θ^U, a removal of ISDS can benefit the host country by bringing implemented investment protection $G(\hat{\theta})$ closer to θ^U.

Proposition 1 *Consider an initial investment agreement that stipulates investment protection $\hat{\theta} \in (\theta^0, \bar{\theta}]$, compensation $T(k, \theta, \hat{\theta})$, and features both ISDS and SSDS. Assume that the stipulated investment protection cannot be implemented with SSDS ($\hat{\theta} \in A^c$). An exclusion of ISDS will then reduce the level of implemented investment protection, to the possible benefit of the host country, and to the disadvantage of the source country.*

This finding thus shows how an exclusion of ISDS might benefit a host country by *eroding implemented protection*, thereby allowing the host country to regulate more often. This is achieved since the political costs facing the source country from enforcing the agreement under SSDS allow the host country to regulate opportunistically in certain instances. Proposition 1 thus gives some support to the notion in the debate that host country should seek to remove ISDS from their agreements, since this might lead to reduced "regulatory chill" from a host country perspective. But it comes at the cost of reduced investment, and it also exposes the host country to political costs for opportunistic violations of the agreement.

Neither partner to an agreement has the power to make such a change to the agreement unilaterally, however. Could the host country induce the source country to accept an exclusion of ISDS in a negotiation? It is indeed possible that the bargaining surplus is larger with SSDS than ISDS, i.e., that $B(\hat{\theta}, S) > B(\hat{\theta}, I)$.[21] But this does not suffice for an exclusion of ISDS to be negotiated, since the source country will lose from such a negotiation. There must by a means through which the host country can compensate the source country for the exclusion of ISDS to be negotiated. We summarize these findings as:

Proposition 2 *Consider an initial investment agreement that stipulates investment protection $\hat{\theta} \in (\theta^0, \bar{\theta}]$, compensation $T(k, \theta, \hat{\theta})$, and features both ISDS and SSDS. Assume that the stipulated investment protection cannot be implemented with SSDS ($\hat{\theta} \in A^c$). The two parties will agree to exclude ISDS in a renegotiation of dispute settlement alone if:*
(i) The host country expected welfare gain from excluding ISDS is sufficiently larger than the source country's expected loss that $B(\hat{\theta}, S) > B(\hat{\theta}, I)$.
(ii) There is a means through which the host country can compensate the source country for the latter's loss from the exclusion of ISDS.

Some of the results in Proposition 2 resemble findings in Ossa, Staiger and Sykes (2020). They develop a model in which the host country has an incentive to expropriate a sunk investment by

---

[21] This property holds for $\hat{\theta} \in A^c$, for instance, if (i) $M^*(\theta, \hat{\theta}) = M(\theta, \hat{\theta}) = 0$, and (ii) $G(\hat{\theta}) = \theta^I$. In this case,

$$B(\hat{\theta}, S) = B(G(\hat{\theta}), S) = B(\theta^I, I) \geq B(\hat{\theta}, I).$$

where the first equality follows from (i), the second equality from (ii), and the inequality from the definition of θ^I. The latter is strict for $\hat{\theta} \neq \theta^I$.

a foreign firm because the arbitration court cannot ascertain whether a taking was efficient or not. This flaw in the dispute settlement regime also creates an incentive for the source country firm/government to litigate in the aspiration to overturn an efficient decision by the host country. Such mistakes are less likely to occur under SSDS relative to ISDS because of less frequent litigation in the former case. The decision whether to implement ISDS or SSDS is taken to maximize joint welfare across the two countries. Whether ISDS is better or worse than SSDS, depends in their model on the quality of the arbitration court, the possibilities for the host country to behave opportunistically, and other factors that determine the implemented investment protection. Hence, they establish explicit circumstances under which $B(\hat{\theta}, S) > B(\hat{\theta}, I)$. Their model satisfies (ii) in Proposition 2 because the host country uses an investment subsidy to compensate the investor for any losses arising from the choice of dispute settlement.

There are also fundamental differences between Proposition 2 and the findings in Ossa, Staiger and Sykes (2020). In their paper, the choice between ISDS and SSDS trades off the benefit of reducing inefficient expropriation against the cost of exacerbating inefficient underregulation. In our model, there is never inefficient underregulation. The choice between ISDS and SSDS instead balances the benefit of reducing inefficient overregulation against the cost of exacerbating a problem of distorted investments emanating from excessive stipulated investment protection.

6 Negotiating both investment protection and dispute settlement

The previous section considered the impact of an exogenous exclusion of ISDS for an arbitrary and constant level of investment protection, as a means of throwing some light on the claim in the policy debate that host countries would benefit from such a measure. However, the degree of investment protection and the form of dispute settlement are both *negotiated* features of the agreements. We therefore in what follows consider the outcome of negotiations regarding both these features. To this end we introduce the bargaining format in the next section, and then separately consider negotiations over the level of investment protection with ISDS and with SSDS separately, before turning to the choice of the form of dispute settlement.

Intuitively, any negotiation in which the two parties also bargain over investment protection must stipulate investment protection $\hat{\theta}$ in excess of θ^0 to fulfill any purpose, for any dispute settlement system. The agreement will simply implement (θ^0, k^0) otherwise (i.e., for all $\hat{\theta} \in [0, \theta^0]$). We formally demonstrate this result in the Appendix. The property $\hat{\theta} > \Theta(K(\hat{\theta}))$, see Lemma 1, implies that there will then be a range of shocks $\theta \in (\Theta(K(\hat{\theta})), \hat{\theta})$ for which the host country would prefer to regulate without compensation, but where the agreement stipulates compensation payments.

6.1 Negotiated protection with ISDS

Assume that the level of investment protection θ^I that maximizes the bargaining function $B(\hat{\theta}, I)$ over $\hat{\theta} \in (\theta^0, \bar{\theta}]$ subject to (11), is unique. The bargained outcome satisfies the necessary first-order condition

$$B_{\hat{\theta}} = \hat{B}_{\tilde{W}} \tilde{W}_{\hat{\theta}} + \hat{B}_{\tilde{W}^*} \tilde{\Pi}_{\hat{\theta}} = 0 \tag{15}$$

for an equilibrium agreement that features incomplete investment protection $\theta^I < \bar{\theta}$, and where both countries strictly benefit from the agreement, with all functions evaluated at $(\hat{\theta}, I)$.

Note that the negotiated outcome reflects a conflict of interest. The marginal effect $\tilde{\Pi}_{\hat{\theta}}$ of investment protection on expected net profit is always positive by (3), so any increase in the level of protection above θ^I would strictly benefit the source country, whereas the country would lose from any reduction below θ^I. It then follows from (15) that $\tilde{W}_{\hat{\theta}}(\theta^I, I) < 0$, so the host country would *benefit from a marginal reduction* in investment protection below θ^I, and it would lose from *any* increase in investment protection above θ^I.[22] By implication, the negotiated investment protection exceeds the level $\theta^U > \theta^0$ that maximizes the host country expected welfare. The inequality is strict if $\theta^U < \bar{\theta}$, which is the case of interest. Finally, the host country would lose for sufficiently large reductions in the level of protection by $\tilde{W}(\theta^I, I) > \tilde{w}^0$.

Lemma 5 (Negotiated protection with ISDS) *Assume that the parties negotiate the level of investment protection given ISDS:*
(i) The negotiation will yield the stipulated level of protection θ^I.
(ii) The agreement will implement θ^I and investment $k^I \equiv K(\theta^I)$.
(iii) The host country would prefer less and the host country more investment protection than θ^I.

6.2 Negotiated protection with SSDS

Negotiation over the level of protection under SSDS results in a stipulated level of protection that maximizes $B(\hat{\theta}, S)$ over $\hat{\theta} \in (\theta^0, \bar{\theta}]$ subject to (11). The analysis now has to account for the possibility that the agreement cannot implement $\hat{\theta}$. Suppose $\hat{\theta} \in A^c \neq \emptyset$, and consider an alternative agreement that stipulates investment protection $G(\hat{\theta})$. The investment $K(G(\hat{\theta}))$ then implements the stipulated protection level $G(\hat{\theta})$ because the combination of investment and stipulated protection satisfies either the source country IC constraint,

$$\Pi(K(G(\hat{\theta}))) = N^*(G(\hat{\theta}), \hat{\theta}), \tag{16}$$

---

[22] For suppose that $\tilde{W}(\hat{\theta}, I) \geq \tilde{W}(\theta^I, I)$ for some $\hat{\theta} > \theta^I$. We would then have $\hat{B}(\tilde{W}(\hat{\theta}, I), \tilde{W}^*(\hat{\theta}, I)) > \hat{B}(\tilde{W}(\theta^I, I), \tilde{W}^*(\theta^I, I))$ by the assumption that the bargaining function is strictly increasing in both arguments, which would violate the presumed optimality of θ^I.

or one of the host country IC constraints:

$$M(G(\hat{\theta}), \hat{\theta}) = \min\{\Pi(K(G(\hat{\theta}))); -V(K(G(\hat{\theta})), G(\hat{\theta}))\}. \tag{17}$$

Hence, $G(\hat{\theta}) \in A$. This alternative agreement implements the same level of investment and protection as the agreement stipulating $\hat{\theta} \in A^c$, but without imposing any political costs on the two countries. It therefore strictly outperforms the initial agreement:

$$B(G(\hat{\theta}), S) = B(G(\hat{\theta}), I) > B(\hat{\theta}, S).$$

We conclude that for every proposed agreement with stipulated protection $\hat{\theta} \in A^c$, there exists an agreement with stipulated protection $G(\hat{\theta}) \in A$ that is strictly better than the proposed agreement. All negotiated levels of investment protection $\hat{\theta}$ under SSDS will therefore be contained in A. Consequently, no agreement negotiated under SSDS generates any political costs for either party in equilibrium:

Lemma 6 (Negotiated protection with SSDS) *Assume that the parties negotiate the level of investment protection given SSDS:*
(i) The negotiation will yield the stipulated level of protection

$$\theta^S \equiv \begin{cases} \theta^I & \text{if } \theta^I \in A \\ \arg\max_{\hat{\theta} \in A} B(\hat{\theta}, I) & \text{if } \theta^I \in A^c. \end{cases}$$

(ii) The agreement will implement θ^S and investment $k^S \equiv K(\theta^S)$.
(iii) There will not be any opportunistic regulation, and hence no political costs, in equilibrium.

If the negotiated outcome with ISDS can be implemented also under SSDS, $\theta^I \in A$, then the negotiating parties will choose this same level also with SSDS. However, if θ^I cannot be implemented under SSDS, $\theta^I \in A^c$, the parties will agree on the implementable level of protection in A that maximizes the bargaining function $B(\hat{\theta}, I)$. Whether $\theta^S \gtrless \theta^I$ is ambiguous, however:

Observation 2 *An agreement with SSDS can feature more or less investment protection than an agreement with ISDS depending on whether θ^I is closer to the upper or lower bound of A^c.*

6.3   Simultaneous negotiation

As emphasized above, investment agreements almost invariably include both ISDS and SSDS. To highlight the forces that work toward the inclusion of ISDS, we now consider a *simultaneous* negotiation over the form of dispute settlement and the stipulated level of protection. This bargaining format corresponds well with how actual negotiations usually are conducted. In our setting, the

25

negotiation maximizes $B(\hat{\theta}, \delta)$ over $\hat{\theta} \in (\theta^0, \bar{\theta}]$ and $\delta \in \{I, S\}$ subject to the contracting parties' participation constraints and other relevant restrictions on the contracting space.

Drawing on Lemmas 5 and 6 we can formally characterize the solution by first solving for the optimal investment protection given ISDS or SSDS separately, and then optimize dispute settlement:

$$
\begin{aligned}
\max_{\hat{\theta} \in (\theta^0, \bar{\theta}]} B(\hat{\theta}, I) &= B(\theta^I, I) \\
&\geq B(\theta^S, I) \\
&= \max_{\hat{\theta} \in (\theta^0, \bar{\theta}]} B(\hat{\theta}, S)
\end{aligned}
$$

The inequality is strict if $\theta^I \in A^c$ because then $\theta^I \neq \theta^S$, whereas θ^I uniquely maximizes $B(\hat{\theta}, I)$. Put differently, the negotiating parties choose from the restricted set A of incentive compatible protection levels under SSDS, whereas they choose from the full set $(\theta^0, \bar{\theta}]$ under ISDS. Consequently:

Proposition 3 (Negotiated dispute settlement) *Simultaneous negotiation over the stipulated investment protection and the dispute settlement mechanism yields an agreement with investment protection θ^I and ISDS.*

This result establishes a first formal rationale for the prevalence of ISDS in investment agreements.

To see why negotiations over both the form of dispute settlement and the level of investment protection will always lead to ISDS, assume that the parties consider some $\hat{\theta} \in A^c$ such that the conditions in Proposition 2 are fulfilled, that is, $\hat{\theta}$ such that a negotiation over only the form of dispute settlement would result in SSDS. Now allow the parties to also negotiate the level of protection. We know from Section 5.2 that if $\hat{\theta} \in A^c$, with SSDS the two countries will agree on investment protection $\theta^S \in A$, since the agreement will then be enforced without any associated political costs. But the agreement $(\hat{\theta}, SSDS)$ can be changed to $(\theta^S, ISDS)$ without affecting the outcome. And we know from Section 5.1 that with ISDS the bargaining will result in the level of protection θ^I, which will generate an even bigger bargaining surplus than θ^S without any drawbacks for either side. Investment protection θ^I sustained by ISDS should therefore be no more difficult to implement than θ^S sustained by ISDS in a revised agreement. That is, the benefits from excluding ISDS that were identified in Proposition 2 rely fundamentally on the assumption that the *stipulated investment protection is unaffected.* This observation has a significant policy implication:

Corollary 1 *Removing ISDS from an agreement is an inefficient solution to problems associated with excessive investment protection. A better solution is to renegotiate investment protection directly under maintained ISDS.*

Proposition 3 does not go all the way to explain why most investment agreements feature both ISDS and SSDS, strictly speaking. The inclusion of SSDS in actual agreements can plausibly be

26

explained by factors falling outside the scope of this model; for instance, it is a standard feature of almost any contract that each party to the contract has the right to question the implementation of the contract by the other party(-ies). The existence of two forms of dispute settlement could in theory lead to a coordination problem in the model in instances where not only the investor, but also the source country government, would be willing to enforce the agreement. There is no point for the firm to enforce compensation payments if the source country does so on its behalf, and vice versa. However, in the present setting where enforcement is without cost for the investor under ISDS, it is always a weakly dominant strategy for the investor to enforce the agreement regardless of the source country government's plan to enforce the agreement. And the political dispute costs for the source country government should help solve this problem. Since a rational host country that anticipates investor enforcement will not litigate, Proposition 3 would then apply even if agreements always contain SSDS by default.

6.4   Negotiating investment protection prior to dispute settlement

As mentioned above, some major recent investment agreement initially only allow for SSDS, but also stipulate that negotiation regarding inclusion of ISDS is to occur at a later stage. Is there any reason to believe that the two-stage negotiation process will make ISDS less likely to be part of the final outcome?

We assume that both stages are negotiated prior to investment and that the objective functions remain the same across the two stages. The second-stage decision about dispute settlement does not matter if the parties have negotiated investment protection $\hat{\theta} \in A$ in the first stage because the agreement will implement $\hat{\theta}$ and investment $K(\hat{\theta})$ regardless of whether ISDS is included in the second round negotiation.

If $\hat{\theta} \in A^c$, then inclusion of ISDS in the second stage will increase the implemented level of investment protection from $G(\hat{\theta})$ to $\hat{\theta}$ and investment from $K(G(\hat{\theta}))$ to $K(\hat{\theta})$. Moreover, the political costs of sustaining investment protection vanish because there is no opportunistic regulation under ISDS. The net effect on the source country and host country expected welfare are given by the negative of the expressions in (12) and (13). Including ISDS in the second stage unambiguously benefits the source country, $\tilde{W}^*(\hat{\theta}, I) > \tilde{W}^*(\hat{\theta}, S)$, and may benefit or hurt the host country. If $\tilde{W}(\hat{\theta}, I) \geq \tilde{W}(\hat{\theta}, S)$, then including ISDS in the second stage represents a Pareto improvement over the first-stage outcome. The condition $\tilde{W}(\hat{\theta}, I) \geq \tilde{W}(\hat{\theta}, S)$ also represents an *interim participation constraint* under sequential bargaining. Unless the source country can compensate the host country for its losses, the agreement will maintain SSDS as the only dispute settlement mechanism, if first-stage investment protection $\hat{\theta}$ implies $\tilde{W}(\hat{\theta}, I) < \tilde{W}(\hat{\theta}, S)$. This result holds even if the bargaining surplus under ISDS is larger than the bargaining surplus under SSDS, so that $B(\hat{\theta}, I) > B(\hat{\theta}, S)$. Based on these results, it would seem as if a sequential bargaining structure increases the possibility for maintaining SSDS as the exclusive dispute settlement mechanism compared to simultaneous

negotiation of investment protection and dispute settlement.

Consider now the first stage negotiation of investment protection under the assumption that the source country cannot compensate the host country for introducing ISDS at the second stage. We can partition A^c into two subsets. The first subset A^c_+ contains all $\hat{\theta} \in A^c$ that satisfy $\tilde{W}(\hat{\theta}, I) \geq \tilde{W}(\hat{\theta}, S)$. The second subset A^c_- contains all $\hat{\theta} \in A^c$ for which $\tilde{W}(\hat{\theta}, I) < \tilde{W}(\hat{\theta}, S)$. By these definitions, the agreement implements any stipulated $\hat{\theta} \in A \cup A^c_+$. In particular, the agreement implements $\hat{\theta} \in A^c_+$ by augmenting dispute settlement with an ISDS provision at the second stage of negotiations. Would the two parties ever negotiate $\hat{\theta} \in A^c_-$? No, because both parties can do strictly better by negotiating $G(\hat{\theta}) \in A$. We therefore conclude:

Proposition 4 *Assume that the parties negotiate the level of investment protection given SSDS in a first stage, and in a second stage whether to include ISDS. Assume also that the two parties do not have access to any side payments.*

(i)   The negotiation will yield the stipulated level of protection

$$\theta^E \equiv \begin{cases} \theta^I & \text{if } \theta^I \in A \cup A^c_+ \\ \arg\max_{\hat{\theta} \in A \cup A^c_+} B(\hat{\theta}, I) & \text{if } \theta^I \in A^c_-. \end{cases}$$

(ii)   The agreement will implement θ^E and investment $k^E \equiv K(\theta^E)$.

(iii) There will be no opportunistic regulation, and hence no political costs, in equilibrium.

Altering the structure of negotiations can have a real effect on the bargaining outcome by changing the equilibrium investment protection from θ^I to θ^E in certain circumstances. Sequential negotiations cannot implement $\theta^I \in A^c_-$ because the host country will veto ISDS in the second stage. The source country is better off under sequential than simultaneous bargaining if $\theta^E > \theta^I$, and the host country is relatively better off under sequential bargaining if $\theta^E < \theta^I$. Hence, the two countries will generally disagree on the bargaining structure, but it is unclear which country prefers which structure because $\theta^E \lessgtr \theta^I$. However, sequential bargaining is less appealing for both parties than simultaneous bargaining in the sense that the former structure is associated with a smaller bargaining surplus: $B(\theta^E, I) < B(\hat{\theta}, I)$. This property pulls in favor of simultaneous bargaining as the most common format. Finally, $\theta^E \in A \cup A^c_+$ implies that the two countries either are indifferent $(\theta^E \in A)$, or strictly prefer $(\theta^E \in A^c_+)$ to include ISDS at the second stage. In particular:

Corollary 2 *Sequential bargaining over investment protection and the form of dispute settlement cannot lead to the exclusion of ISDS in equilibrium.*

This economic logic suggests that the recent large agreements with sequential negotiations of dispute settlement will ultimately include ISDS. However, this will require that the governments involved can overcome the current popular opposition to ISDS, which is far from obvious they will do.

7 Concluding remarks

The policy debate on investment agreements centres on the appropriate scope of their substantive rules and the design of the dispute settlement mechanisms that enforce these provisions. A common claim in the debate is that host countries would be better off without ISDS. We show that excluding ISDS from an agreement can indeed increase the host country's ability to regulate without facing litigation if the source country faces political costs of enforcing an agreement, for a given stipulated level of investment protection. Such costs will reduce investment protection under SSDS relative to ISDS. Excluding ISDS from an agreement that for some reason stipulates excessive investment protection from the point of view of the a host country can therefore be one way to increase its regulatory space. Exclusion of ISDS comes with costs however, since the reduced level of investment protection reduces investment, and since SSDS can generate political costs by creating incentives for host country opportunistic regulation. These costs must be weighted against any benefits of reduced protection. But the host country might still on balance benefit from the exclusion of ISDS. It is even possible that the host country could negotiate an exclusion of ISDS in certain cases, provided that there is a means of compensating the source country for its resulting loss.

These findings have assumed a given level of investment protection. However, the parties negotiate both the level of investment protection and the form of dispute settlement when they form agreements. Ours is the first integrated economic analysis of these two issues. It is based on the standard explanation in the legal literature for the inclusion of ISDS and SSDS in invest agreements—the desire to avoid the conflicts that state-state disputes can give rise to. We demonstrate that when two countries negotiate dispute settlement in an international investment agreement, under general conditions they will include the possibility for investors to litigate against the host country. Hence, a more efficient remedy than removal of ISDS, would be to renegotiate the stipulated investment protection directly so as to allow additional carve-outs from compensation requirements. This observation is consistent with the fact that most of the more recent investment agreements still allow for ISDS in some form, but also include explicit carve-out provisions.

Finally, a couple of major recent investment agreement include SSDS from the outset, but postpone negotiations regarding the inclusion of ISDS to the future. We demonstrate why such a bargaining procedure is likely to ultimately result in the inclusion of ISDS, as in the case of a simultaneous negotiation. But that it might still affect the negotiated level of protection.

We conclude by discussing three directions for research on dispute settlement in investment agreements.

Rationales for SSDS In our model, SSDS constrains the outcome relative to what parties can negotiate under ISDS, depending on the magnitude of the political costs associated with SSDS. Providing a strictly beneficial role for SSDS under joint negotiation of dispute settlement and investment protection would require an extended framework. Such a framework could entail financial

constraints that prevent private investors from initiating disputes. Source country governments can also find it more worthwhile than firms to pursue particular types of disputes. For instance, formal settlement of a dispute can contribute to the case law, which increases the efficiency of future dispute settlement. Such spill-over effects would be particularly likely to occur in case of policy measures that affect multiple firms in an industry. Similarly, fixed costs for legal processes might prevent individual investors from pursuing disputes that would be profitable at an industry level. The source country government could then act as a substitute for investor class action. Including both ISDS and SSDS in an agreement then increases the source country freedom to pick the least-cost enforcement of the agreement, which in turn increases the bargaining surplus compared to an agreement that relies exclusively on ISDS or SSDS.

Equilibrium disputes   The critique against investment agreements has been fuelled by a number of high-profile investment disputes, such as the litigation by the energy company Vattenfall against Germany for the phase-out of commercial nuclear power in the wake of the Fukushima disaster, the $15 billion lawsuit against the US by TransCanada Corporation for the rejection of the Keystone pipeline, and the Phillip Morris litigation against Australia for the introduction of tobacco plain packaging legislation. Furthermore, with more than 1 100 known disputes under investment treaties, almost all of which ISDS, investment disputes can said to be very common.

Our framework does not yield disputes in equilibrium. While it would be desirable to be able to capture also this aspect of investment agreements, this apparent weakness should be set against several features of investment disputes and of the present framework.

First, the number of recorded disputes is small compared to the number of agreements. Currently, there are more than 2 500 agreements in force. Aggregated over time, there has been less than half a dispute per agreement on average. In addition, there seems to be a rather high degree of concentration of disputes to a limited number of agreements; for instance, the Energy Charter Treaty alone accounts for more than 140 disputes.

Second and more importantly, the main focus of the analysis is on the role of dispute settlement system when it is jointly negotiated with the level of investment protection. The model does capture how the *threat* of disputes and the associated political costs affect these variables, as well as the regulatory space of the host country.

Third, it is a very hard task to provide an empirically plausible and reasonably general explanation of why investment disputes arise. These difficulties are reflected in the fact that the vast theory literature on the parallel form of treaties—trade agreements—hardly include any models that feature disputes in equilibrium, despite the interest that the WTO dispute settlement system attracts in the policy debate and in research, and despite the more than 600 disputes that have arisen under this trade agreement alone. Multi-sided informational asymmetry is a general source of inefficiency in contractual relationships, which in the present context would occur as arbitration through ISDS or SSDS. It also seems plausible that host country governments may engage in disputes for domestic

political reasons, or as a strategy to deter future litigation. Similar mechanisms have been analyzed in incomplete information games.

But the more basic reason for disputes seems to be the *vague drafting* of substantive obligations in agreements. Were it not for this vagueness, there would hardly be the expanding professions of trade and investment lawyers, since the outcome of disputes would essentially be known beforehand. We simply lack appropriate micro-economic tools to model this form of contractual incompleteness of agreements.

Some models of investment agreements (e.g. Aisbett, Karp and McAusland, 2010a; Ossa, Staiger and Sykes, 2020) do feature equilibrium disputes. These occur by an assumption that the arbitration tribunal cannot perfectly monitor the state of the world.[23] This deficiency creates an incentive for the host country to regulate without paying required compensation and for the investor (or source country under SSDS) to litigate in order to correct this injustice. Excessive litigation can also occur in the aspiration to obtain undue compensation payments. However, monitoring and verification problems alone are insufficient to generate equilibrium disputes if the host country and the investor (or source country under SSDS) can settle prior to adjudication. The costs and the risks involved with arbitration imply that there are gains from settling without invoking a tribunal. Indeed, many investment agreements contain provisions that encourage and enable amicable dispute settlement. It is also very common in practice that disputes end with a settlement, sometimes through mediation, rather than arbitration. The models that generate equilibrium arbitration achieve this by assuming that compensation can only be implemented through formal dispute settlement. The present model in contrast allows the host country to abide by the rules in the agreements regarding compensation, precisely to avoid a costly legal process.

Investment protection is insensitive to industry-specific circumstances   We have considered an agreement that is tailor-made for one particular industry. But investment agreements typically have economy-wide application with few industry-specific obligations. To illustrate potential implications of such rigidity, assume that source country firms can invest in two economically unrelated industries in the home country. The investment agreement specifies a single level of protection $\hat{\theta}$ that applies to both industries. Assume for simplicity that the source country is the dominant party in the negotiations and can set $\hat{\theta}$ unilaterally. It then chooses investment protection such that the host country is indifferent between accepting and rejecting the agreement. The host country must then lose from the inclusion of one of the sectors in the agreement if the two

---

[23]It is straightforward to see that if the outcome of arbitration regarding an unlawful regulation were assumed to be stochastic in our setting, there could be an incentive for a host country to take such a measure also when this will lead to a dispute. For instance, this might arise in case a dispute leds to the requested payment

$$T(k,\theta,\hat{\theta}) \equiv \begin{cases} \Pi(k) & \text{with probability } \omega \text{ if } \theta \leq \hat{\theta} \\ 0 & \text{with probability } 1-\omega \text{ if } \theta \leq \hat{\theta}, \text{ or if } \theta > \hat{\theta} \end{cases}$$

It would be natural to assume that ω is declining in the magnitude of the regulatory shock, and increasing in the negotiated level of protection. In the above we have used a special case of this payment function, where $\omega = 1$.

industries differ in the welfare they generate for the host country. Denote the losing industry by z. If the political enforcement costs under SSDS are sufficiently high that the source country will never enforce any investment protection in industry z, investors will no longer expect protection in this sector if ISDS is excluded. They will therefore invest in this industry as if there was no agreement. Excluding ISDS with unchanged investment protection can therefore enable the host country to escape harmful protection commitments in certain industries, when obligations are not fully adapted to industry-specific conditions.

This is still not a very strong argument in favor of SSDS. There is no guarantee that the unravelling of the agreement occurs only in industries where the host country would prefer there to be no agreement. Also, exclusion of ISDS might plausibly lead to a renegotiation of the level of investment protection. However, these mechanisms do illustrate why it might be interesting to address questions related to dispute settlement in agreements that lack industry-specific obligations.

A   Appendix

A.1   Proof of Lemma 1

If $\theta' - \Theta(K(\theta')) \leq 0$ for some $\theta' \in (\theta^0, \bar{\theta}]$, then there exists $\theta_t^0 \in [\theta', \bar{\theta}]$ such that $\theta_t^0 - \Theta(K(\theta_t^0)) = 0$ by $\bar{\theta} - \Theta(K(\bar{\theta})) \geq 0$ and an application of the Intermediate Value Theorem. The pair (k_t^0, θ_t^0), where $k_t^0 = K(\theta_t^0)$, constitutes a Nash equilibrium of the game when there is no investment agreement. Since $\theta_t^0 \geq \theta' > \theta^0$, this contradicts the assertion that (k^0, θ^0) is the maximal equilibrium. Hence, $\theta' - \Theta(K(\theta')) > 0$ for all $\theta' \in (\theta^0, \bar{\theta}]$.∎

A.2   Proof of Lemma 4

Suppose the firm has invested $K(\theta')$ in the expectation that the agreement will enforce investment protection $\theta' \leq \hat{\theta}$ under SSDS. Dispute settlement only matters for the implemented investment protection if $\Theta(K(\theta')) < \hat{\theta}$ and in that case only for shocks $\theta \in (\Theta(K(\theta')), \hat{\theta}] \equiv \Lambda(\theta', \hat{\theta})$. The host country will always allow production for shocks $\theta \leq \Theta(K(\theta'))$ even absent any agreement and legally regulate without compensation for all $\theta > \max\{\hat{\theta}; \Theta(K(\theta'))\}$. We first characterize the implemented investment protection $L(\theta', \hat{\theta})$.

Denote by $L^*(\theta', \hat{\theta})$ the implemented investment protection offered through *source country enforcement* in an agreement with stipulated investment protection $\hat{\theta}$ and investment $K(\theta')$ if this is the only way to enforce the agreement. For $\theta \in \Lambda(\theta', \hat{\theta})$, the investor will recover its operating profit by the threat of litigation through SSDS if:

$$\Pi(K(\theta')) \geq N^*(\theta, \hat{\theta}).$$

By monotonicity of $N^*(\theta, \hat{\theta})$ in θ, $L^*(\theta', \hat{\theta}) \equiv \hat{\theta}$ if $N^*(\hat{\theta}, \hat{\theta}) \leq \Pi(K(\theta'))$ and $L^*(\theta', \hat{\theta}) \equiv \Theta(K(\theta'))$

if $N^*(\Theta(K(\theta')), \hat{\theta}) \geq \Pi(K(\theta'))$. For intermediary enforcement cost, the source country will only enforce the agreement for sufficiently mild shocks, $\theta \in (\Theta(K(\theta')), L^*(\theta', \hat{\theta})]$, where

$$\Pi(K(\theta')) \equiv N^*(L^*, \hat{\theta})$$

defines $L^*(\theta', \hat{\theta}) \in (\Theta(K(\theta')), \hat{\theta})$.

Denote by $L^1(\theta', \hat{\theta})$ the actual investment protection offered through the *host country's prefer-ence for regulation with immediate compensation relative to engaging in opportunistic regulation*, given stipulated investment protection $\hat{\theta}$ and investment $K(\theta')$, if this is the only way to enforce the agreement. The investor can rely on the host country to honor the agreement for $\theta \in \Lambda(\theta', \hat{\theta})$ if

$$M(\theta, \hat{\theta}) \geq \Pi(K(\theta')).$$

By monotonicity of $M(\theta, \hat{\theta})$ in θ, $L^1(\theta', \hat{\theta}) \equiv \hat{\theta}$ if $M(\hat{\theta}, \hat{\theta}) \geq \Pi(K(\theta'))$ and $L^1(\theta', \hat{\theta}) \equiv \Theta(K(\theta'))$ if $M(\Theta(K(\theta')), \hat{\theta}) \leq \Pi(K(\theta'))$. For intermediary political cost, the host country honors the terms of the agreement if $\theta \in (\Theta(K(\theta')), L^1(\theta', \hat{\theta})]$, where

$$\Pi(K(\theta')) \equiv M(L^1, \hat{\theta})$$

defines $L^1(\theta', \hat{\theta}) \in (\Theta(K(\theta')), \hat{\theta})$.

Denote by $L^2(\theta', \hat{\theta})$ the actual investment protection offered through the *host country's pref-erence for production relative to engaging in opportunistic regulation*, given stipulated investment protection $\hat{\theta}$ and investment $K(\theta')$, if this is the only way to enforce the agreement. For $\theta \in \Lambda(\theta', \hat{\theta})$, the host country prefers production to opportunistic regulation if

$$V(K(\theta'), \theta) \geq -M(\theta, \hat{\theta}).$$

By monotonicity of $V(k, \theta)$ and $M(\theta, \hat{\theta})$ in θ, $L^2(\theta', \hat{\theta}) \equiv \hat{\theta}$ if $V(K(\theta'), \hat{\theta}) + M(\hat{\theta}, \hat{\theta}) \geq 0$. If the value of production is small, $V(K(\theta'), \hat{\theta}) + M(\hat{\theta}, \hat{\theta}) < 0$, then the host country allows production for all shocks $\theta \in (\Theta(K(\theta')), L^2(\theta', \hat{\theta})]$, where $L^2(\theta', \hat{\theta}) \in (\Theta(K(\theta')), \hat{\theta})$ is defined by:

$$V(K(\theta'), L^2) \equiv -M(L^2, \hat{\theta}).$$

The actual investment protection implemented through the source country and host country incentive compatibility constraints under SSDS is given by

$$L(\theta', \hat{\theta}) \equiv \max\{L^*(\theta', \hat{\theta}); L^1(\theta', \hat{\theta}); L^2(\theta', \hat{\theta})\}.$$

in an agreement with stipulated investment protection $\hat{\theta}$ where the firm has invested $K(\theta')$ in the expectation of receiving investment protection $\theta' \leq \hat{\theta}$. The function $L(\theta', \hat{\theta})$ is continuous by

continuity of $V(k, \theta)$, $\Pi(k)$ and $K(\theta')$ and by the properties of $N^*(\theta, \hat{\theta})$ and $M(\theta, \hat{\theta})$. If the firm has invested $K(\hat{\theta})$ and either (6), (7) or (8) holds so that $\hat{\theta} \in A$, then the configuration of political costs implements stipulated investment protection $\hat{\theta}$. The firm's expectation of investment protection $\hat{\theta}$ is confirmed in this case: $L(\hat{\theta}, \hat{\theta}) = \hat{\theta}$.

If the source country enforcement cost is large, $N^*(\hat{\theta}, \hat{\theta}) > \Pi(K(\hat{\theta}))$ and the host country political cost of opportunistic regulation is small, $M(\hat{\theta}, \hat{\theta}) < \min\{\Pi(K(\hat{\theta})); -V(K(\hat{\theta}), \hat{\theta})\}$, so that $\hat{\theta} \in A^c$, then a firm that has invested $K(\hat{\theta})$ will receive less investment protection than stipulated in the agreement because of opportunistic regulation by the host country: $\hat{\theta} - L(\hat{\theta}, \hat{\theta}) > 0$. Suppose instead the investor is very pessimistic and anticipates that the agreement will offer no investment protection beyond what the investor would receive without any agreement, $\theta' = \theta^0$, and invests accordingly, $K(\theta') = k^0$. In this case

$$V(k^0, \theta^0) + M(\theta^0, \hat{\theta}) = M(\theta^0, \hat{\theta}) > 0$$

implies $L^2(\theta^0, \hat{\theta}) > \theta^0$ and therefore $\theta^0 - L(\theta^0, \hat{\theta}) < 0$. This pessimistic belief understates actual investment protection.

By continuity of $L(\theta', \hat{\theta})$ in θ', we can apply the Intermediate Value Theorem to establish existence of $\theta' \in (\theta^0, \hat{\theta})$ with the property $\theta' = L(\theta', \hat{\theta})$. For all such θ' a firm that invests $K(\theta')$ will receive investment protection θ'. There can be multiple such θ', so we define $G(\hat{\theta})$ as the maximal solution in $(\theta^0, \hat{\theta})$ that features $\theta' = L(\theta', \hat{\theta})$. Hence, $G(\hat{\theta})$ is *the maximal investment protection that an investor can receive in rational expectations equilibrium when an agreement with SSDS cannot implement $\hat{\theta}$ subject to investment $K(\hat{\theta})$.* Since $G(\hat{\theta}) < \hat{\theta}$, this investment protection solves either (16) or (17).

If the firm has invested $K(G(\hat{\theta}))$, the host country allows production for all $\theta \leq G(\hat{\theta})$ if $\Pi(K(G(\hat{\theta}))) \leq -V(K(G(\hat{\theta})), G(\hat{\theta}))$. In the opposite situation, $\Pi(K(G(\hat{\theta})) > -V(K(G(\hat{\theta})), G(\hat{\theta}))$, the host country allows production for all $\theta \leq \Theta^J(K(G(\hat{\theta}))$ and regulates with immediate compensation for all $\theta \in (\Theta^J(K(G(\hat{\theta})), G(\hat{\theta})]$. Either way, the host country engages in opportunistic regulation for all $\theta \in (G(\hat{\theta}), \hat{\theta}]$ and legally regulates without paying compensation for all $\theta > \hat{\theta}$ if $\hat{\theta} < \bar{\theta}$. The host country expected welfare and the source country expected welfare are then given by (9) and (10).

To verify that the agreement indeed implements $(K(G(\hat{\theta})), G(\hat{\theta}))$, we need to show that the agreement cannot implement any investment protection $\theta' > G(\hat{\theta})$. Suppose the firm expects investment protection $\theta' \in (G(\hat{\theta}), \hat{\theta}]$ and invests $K(\theta')$ accordingly. Since $G(\hat{\theta})$ is defined as the maximal fixed point contained in $(\theta^0, \hat{\theta})$, $\theta' > L(\theta', \hat{\theta})$ for all $\theta' \in (\theta^0, \hat{\theta})$, which is inconsistent with rational expectations. This holds also for $\theta' = \hat{\theta}$ since $\hat{\theta} > L(\hat{\theta}, \hat{\theta})$ by $\hat{\theta} \in A^c$. Suppose the firm expects investment protection $\theta' > \hat{\theta}$, $\hat{\theta} < \bar{\theta}$, and invests $K(\theta') > K(\hat{\theta})$. Since $\theta' > \theta^0$, $\theta' > \Theta(K(\theta'))$ by Lemma 1. Therefore, $V(K(\theta'), \theta) \leq V(K(\theta'), \theta'') < V(K(\theta'), \Theta(K(\theta'))) = 0$ for all $\theta \in [\theta'', \bar{\theta}]$, and some $\theta'' \in (\max\{\Theta(K(\theta')); \hat{\theta}\}, \theta')$. The host country will therefore regulate without paying

compensation for all $\theta \in [\theta'', \bar{\theta}]$. Since $\theta'' < \theta'$, this is again inconsistent with the firm's expectation of investment protection θ'.∎

A.3 A general property of negotiated investment protection

Lemma 7 *Any negotiated investment agreement that stipulates compensation $T(k, \theta, \hat{\theta})$ and dispute settlement $\delta \in \{I, S\}$, has stipulated investment protection $\hat{\theta} \in (\theta^0, \bar{\theta}]$.*

Proof: Suppose $\hat{\theta} \in [0, \theta^0]$. If the firm invests k^0, then the host country prefers production to regulation for all $\theta \leq \theta^0 = \Theta(k^0)$ even if regulation is completely without cost, since $V(k^0, \theta) \geq V(k^0, \theta^0) = 0$ for all $\theta \leq \theta^0$. Conversely, it is optimal to regulate without compensation for all $\theta > \Theta(k^0) \geq \hat{\theta}$. Since k^0 yields the threshold θ^0 for regulation, and k^0 is the profit-maximizing investment given the threshold θ^0, (k^0, θ^0) can be sustained as an equilibrium also under the proposed agreement. As (k^0, θ^0) is the maximal equilibrium, this is the one that will be implemented by way of the assumption regarding equilibrium selection. Since $\tilde{W}(\hat{\theta}, \delta) = \tilde{w}^0$ and $\tilde{W}^*(\hat{\theta}, \delta) = \tilde{w}^{*0}$, no country strictly benefits from an agreement that features $\hat{\theta} \in [0, \theta^0]$. Hence, any investment agreement that strictly increases joint surplus must have stipulated investment protection $\hat{\theta} \in (\theta^0, \bar{\theta}]$.∎

References

Aisbett, Emma, Larry Karp and Carol McAusland (2010a). "Police Powers, Regulatory Takings and the Efficient Compensation of Domestic and Foreign Investors." *The Economic Record* 86(274), September, 367–383.

Aisbett, Emma, Larry Karp and Carol McAusland (2010b). "Compensation for Indirect Expropriation in International Investment Agreements: Implications of National Treatment and Rights to Invest." *Journal of Globalization and Development* 1(2), 1-35.

Bernasconi-Osterwalder, Nathalie (2014). "State-State Dispute Settlement in Investment Treaties". International Institute for Sustainable Development Best Practice Series. October.

Dolzer, Rudolf and Christoph Schreuer (2012). *Principles of International Investment Law.* Oxford, UK: Oxford University Press.

Gertz, Geoffrey (2017). "Renegotiating NAFTA: Options for Investment Protection". *Globalviews* 7, March. Brookings Institution.

Howse, Robert L. (2017). "International Investment Law and Arbitration: A Conceptual Framework." In Helene Ruiz-Fabri (ed.), *International Law and Litigation*, New York: Nomos Press.

Horn, Henrik and Thomas Tangerås (2021). "Economics of International Investment Agreements". *Journal of International Economics* 131, 103433.

Janeba, Eckhard (2019). "Regulatory Chill and the Effects of Investor State Dispute Settlement." *Review of International Economics* 27(4), 1172-1198.

Janeba, Eckhard (2020). "Investor State Dispute Settlement and Host Country Regulation: Insights from Economic Theory." In: Chaisse J., Choukroune L., Jusoh S. (eds) *Handbook of International Investment Law and Policy.* Springer, Singapore.

Kohler, Wilhelm and Frank Stähler (2019). "The Economics of Investor Protection: ISDS versus National Treatment." *Journal of International Economics* 121, 103254.

Konrad, Kai A. (2017). "Large Investors and Permissive Regulation: Why Environmentalists May Dislike Investor-State Dispute Settlement." *European Economic Review* 98, 341-353.

Kreibaum, Ursula (2018). "Evaluating Social Benefits and Costs of Investment Treaties: Depoliticization of Investment Disputes." *ICSID Review* 33(1), 14-28.

Markusen, James R. (1998). "Multilateral Rules on Foreign Direct Investment: the Developing Countries' Stake." World Bank Working Paper.

Markusen, James R. (2001). "Commitment to Rules on Investment: The Developing Countries' Stake," *Review of International Economics* 9(2), 287–302.

Menon, Trishna (2018). "Developing Country Opposition to an Investment Court: Could State-State Dispute Settlement be an Alternative?". Kluwer Arbitration Blog, February 17.

Newcombe, Andrew and Lluis Paradell (2009). *Law and Practice of Investment Treaties.* Kluwer Law International.

Ossa, Ralph, Robert W. Staiger and Alan O. Sykes (2020). "Disputes in International Investment and Trade." NBER Working Paper 27012.

Paparinskis, Martins. (2010). "The limits of depoliticisation in contemporary investor-state arbitration." In: Crawford J, Nouwen S (eds), *Select proceedings of the European Society of International Law,* vol 3. Hart, pp 271–282.

Pohl, J., K. Mashigo and A. Nohen (2012). "Dispute Settlement Provisions in International Investment Agreements: A Large Sample Survey." OECD Working Papers on International Investment No. 2012/02,. OECD Publishing, Paris.

Roberts, Anthea (2014). "State-to-State Investment Treaty Arbitration: A Hybrid Theory of Interdependent Rights and Shared Interpretative Authority". *Harvard International Law Journal 55(1),* 1-70.

Shihata, Ibrahim (1986). "Towards a Greater Depoliticization of Investment Disputes: The Roles of ICSID and MIGA." *ICSID Review - Foreign Investment Law Journal* 1(1), 1-25.

Stiglitz, Joseph E. (2008). "Regulating Multinational Corporations: Towards Principles of Cross-Border Legal Frameworks in a Globalized World Balancing Rights with Responsibilities." 2007 Grotius Lecture. *American University International Law Review* 23, 451-558.

Salacuse, Jeswald W. (2007). "Is There a Better Way? Alternative Methods of Treaty-Based Investor-State Dispute Resolution". *Fordham International Law Journal 31(1)*, 138-185.

Sykes, Alan O. (2005). "Public vs. Private Enforcement of International Law: Of Standing and Remedy". *Journal of Legal Studies 34,* 631.

Trevino, Clovis J. (2013). "State-to-State Investment Treaty Arbitration and the Interplay with Investor-State Arbitration Under the Same Treaty". *Journal of International Dispute Settlement 5*, 199-233.

Turrini, Alessandro and Dieter M. Urban (2008): "A Theoretical Perspective on Multilateral Agreements on Investment." *Review of International Economics* 16(5), 1023–1043.

UNCTAD (2014). Investor-State Dispute Settlement. UNCTAD Series on Issues in International Investment Agrements II.

Vandewelde, Kenneth J. (1988). "The Bilateral Investment Treaty Program of the United States." *Cornell International Law Journal* 21(2), 201-276.