Journal of Economics (2023) 138:1–16
https://doi.org/10.1007/s00712-022-00800-z

An optimal investor-state dispute settlement mechanism

Frank Stähler[1]

Received: 19 January 2021 / Accepted: 27 June 2022 / Published online: 6 August 2022

Abstract
Many investment treaties include investor-state dispute settlement (ISDS) provisions which are supposed to protect a foreign investor against opportunistic behavior of a host country. This paper scrutinizes the optimal design of ISDS provisions that solve the holdup problem. It shows that an efficient investor protection mechanism requires an arbitrator as established in investment treaties. However, this arbitrator does neither have to learn nor to evaluate the circumstances of the dispute. Furthermore, any ISDS compensation from the government to the investor should not be based on reductions in investor profits but on the host country's welfare effects.

Keywords Investor state dispute settlement · Foreign direct investment · TTIP · TPP

JEL Classification F21 · F23 · F53 · F55 · D82 · H13

1 Introduction

Investor-state dispute settlements (ISDS) provisions seem to be the most controversial policy issue in international trade policies in recent times. ISDS provisions are blamed to undermine national sovereignty as they allow foreign investors to take any kind of apparently unfair treatment to a tribunal. This tribunal may rule that the host country of the investment will have to indemnify the foreign investor if they find that the host country government policies have caused "unjustified" harm equivalent to expropriation. While ISDS provisions can be found in many bilateral investment treaties (see OECD 2012), the political opposition against ISDS provisions gained momentum when it became clear that they were supposed to become an integral part

✉ Frank Stähler
frank.staehler@adelaide.edu.au

[1] University of Adelaide, University of Tübingen, CESifo and NoCet, Mohlstr. 36 (V4), 72074 Tübingen, Germany

of both the Transatlantic Trade and Investment Partnership (TTIP) and the Trans-Pacific Partnership (TPP) Agreement.[1]

For example, Article 9.8 of the TPP draft (2016) specified that "[n]o Party shall expropriate or nationalize a covered investment either directly or indirectly through measures equivalent to expropriation or nationalization (expropriation) ..." Regulatory chill is a further concern and means that national governments, anticipating potential compensations, will constrain themselves and may not pursue any policy that might affect future multinational profits, leading to lower regulatory standards in host countries. Although the Trump administration abandoned both TTIP and TPP, ISDS provisions can be expected to become an integral part also in future agreements, both bilateral and multilateral.[2]

The economic reasoning behind ISDS provisions is the well-known holdup problem that may arise in the context of foreign direct investment (FDI). Due to incomplete contracts, a government cannot credibly promise investor-friendly policies if the investment is very specific such that it has little or no value when relocated to another country.[3] As a consequence, some beneficial investment may not take place as the foreign investor will correctly anticipate her exposure to the holdup problem. International investment treaties specify rules by which disputes should be resolved.[4] This paper scrutinizes how these rules should be specified.

There is a substantially large literature that investigates empirically the effects of international investment treaties on foreign direct investment. This literature finds that they increase investment (see Busse et al. 2010; Egger and Merlo 2012; Egger and Pfaffermayr 2004; Haftel 2010; Neumayer and Spess 2005; Rose-Ackerman and Tobin 2009, 2011), but the impact of ISDS provisions on FDI activities is unclear (see Berger et al. 2011). The theoretical literature on ISDS is not large. There is, however, a related literature on compensations for regulatory takings, see Blume et al. (1984), Hermalin (1995), Miceli and Segerson (1994) and Nosal (2001).

As for ISDS, Kohler and Stähler (2019) find that an ISDS provision may improve aggregate welfare generated by a relationship between a foreign investor and a host country if strategic ownership changes could be ruled out. However, the ISDS provision will lead to further distortions and can never achieve the first best. Aisbett et al. (2010a) show that taking out some well-defined policies from any potential compensation claim (so-called "police powers carve-outs") can improve even the host country's welfare. Janeba (2019) discusses under which conditions ISDS can lead to regulatory chill, and he shows that a unilateral domestic ISDS provision

[1] TTIP was negotiated between the US and the EU, and TPP was already ratified by several Pacific rim countries, including the US, Canada, Australia and Japan.

[2] TTP came into force as TTP-11, i.e., without the US, but TPP-11 does not contain any ISDS and intellectual rights provisions that the TPP draft suggested. CETA, an agreement between Canada and the EU, came into force, and it includes an ISDS provision.

[3] See Navaretti and Venables (2006) for an exposition of the standard holdup problem. The allocation of property rights can mitigate the holdup problem as shown by Antràs (2003), Antràs and Helpman (2004) and Antràs and Chor (2013).

[4] These provisions refer to legal rules as they are either set by the United Nations Commission On International Trade Law (UCITRAL) or the International Centre for Settlement of Investment Disputes (ICSID) which is part of the World Bank.

reduces domestic welfare, but a bilateral ISDS provision may (or may not) increase welfare. Konrad (2017) develops a model with strategic interactions between a domestic and a foreign firm, but without an opportunistic government. In his model, the domestic firm will strategically over-invest without ISDS, and an ISDS provision will ensure equal treatment of the domestic and the foreign firm, leading to even more over-investment and overly permissive regulation. Schjelderup and Stähler (2021) show that an ISDS provision in an agreement among institutionally strong countries makes multinational firms more aggressive in terms of investment and market behavior.

Overall, most of the empirical and the sparse theoretical literature has discussed the issue of investor protection only by looking upon the implications of past provisions and the expected effects of ISDS provisions for future multilateral agreements. This paper will not take ISDS provisions as given, but will develop a simple optimal mechanism that is able to solve the holdup problem. Consequently, the innovation of this paper is that it outlines how an ISDS provision should be designed that may not only mitigate the holdup problem, but will solve it. Aisbett et al. (2010b) consider the efficient compensation of (domestic and foreign) investors, but their model assumes a court that receives a stochastic signal on the harmfulness of the investment, and regulation will always shut down the investor's operation. In the model of this paper, no third party receives any signal on the true intervention necessity.

The papers closest to the model of this paper are Aisbett and Bonnitcha (2021), Bonnitcha and Aisbett (2021). These papers develop an optimal compensation rule, too, and their analysis accommodates both opportunistic behavior and new and unanticipated information. Unanticipated information implies that treaties cannot be complete. The difference is that they assume that benefits and costs are common knowledge and can thus be verified. This paper does not deal with unanticipated information, but does not assume that costs and benefits can be verified by any third party.

There is also a substantial literature on arbitration and (re-)negotiation in trade agreements through institutions like the WTO, and most papers also assume in this context that an arbitration panel will receive a stochastic signal [see, for example, Beshkar (2016)].[5] The arbitrator in our model does not receive any signal, but manages an arbitration process according to rules the investor and the government have agreed upon.[6] The paper will show that such a mechanism must involve at least three parties, the investor, the government and an arbitrator. Since the mechanism does not work without arbitrator, the paper emphasizes that a proper investor protection mechanism could and should be managed by a supranational institution.

[5] Furthermore, another difference is that trade agreements may allow complainants to retaliate through (inefficient) trade measures, but they do not include compensations like ISDS provisions do. For this strand of literature, see Beshkar (2010a, 2010b, 2016), Klimenko et al. (2007), Maggi and Staiger (2011, 2015, 2018) and Park (2011).

[6] Horn and Tangerås (2016) show that an optimal investment agreement can be designed such that the investor will be compensated if and only if a regulatory shock is below a specified threshold. In their model, however, the regulator has only the choice to shut down the multinational operation, whereas this paper allows for a range of regulatory activities.

Therefore, the paper will also make a contribution on the role of multilateralism versus bilateralism for investment liberalization.[7]

The implementation literature has dealt with holdup problems in different contexts, see for example Moore and Repullo (1988) and Maskin and Tirole (1999). The seminal models have been developed in the context of incomplete contracts between private agents like firms and not between a government and a firm. The innovation of this paper is (1) to extend this analysis to ISDS and (2) to suggest a simple design that guarantees continued investor activity also off the equilibrium path. The implementation literature often suggests the stop of any activity as an alternative action plan, but this paper shows that an optimal ISDS design is feasible without the threat of discontinuation. In particular, all what is needed are two action plans and two transfer schemes, where one action plan has only to be less ambitious than the other.

Consequently, the remainder of the paper is organized as follows. Section 2 will set up a model that will imply the holdup problem in the absence of any investor protection. Section 3 will develop a simple optimal ISDS mechanism, and Sect. 4 will present an example. Section 5 will offer some concluding remarks.

2 The model

The model is an extension of the standard models of foreign direct investment that is specific and potentially subject to opportunistic behavior. The innovation is that the future environment as it affects domestic welfare is *ex ante* uncertain, and that we consider three agents in this model: the domestic government, the foreign investor, and an arbitrator. The arbitrator may become active only if the domestic government and the foreign investor agree on an ISDS provision. In detail, the maximized investor profit after entry, denoted by π, is assumed to depend negatively on the activity level a:

$$\pi(a) : \pi_a(a) < 0, \pi_{aa}(a) \leq 0, \exists \bar{a} > 0 : \pi(\bar{a}) = \bar{\pi} \geq 0, \tag{1}$$

where subscripts denote partial derivatives. The government exercises a regulatory activity after entry, denoted by $a, a \in [0, \infty[$, that will harm the investor.[8] A higher activity level is equivalent to a higher level of intervention or an increase in regulation affecting the investor. The profit level $\bar{\pi}$ gives the value of the investment when leaving the country. According to (1), an activity level $\bar{a}$ exists such that investor profit will fall below $\bar{\pi}$ for all $a > \bar{a}$. If the investment is completely country-specific such that it is of no value at any other location, $\bar{\pi} = 0$, and $a > \bar{a}$ would make investor profit become negative. In any case, the investor will leave the domestic

[7] See for the role of multilateralism versus bilateralism in the context of trade liberalization, Bagwell and Staiger (2005) and Bagwell et al. (2016).

[8] For the sake of simplicity, we assume a to be a scalar, but our results do also hold if it were a vector of activities. Horn and Tangerås (2016), Janeba (2019) and Konrad (2017) consider the case that the government has a binary policy choice, and Horn and Tangerås (2016), Konrad (2017) and Schjelderup and Stähler (2021) allow for variable investment costs. This paper considers flexible regulatory activities, but assumes fixed entry costs.

country if $a > \bar{a}$. Furthermore, entry warrants a fixed cost of market entry, denoted by $\Phi > 0$, that cannot be recovered after entry.

Domestic welfare depends (1) on the activity level itself which benefits the domestic country if it is not excessively large, and (2) on the realization of a stochastic variable θ that measures the degree of intervention necessity.[9] Importantly, this realization becomes known only to the government and the investor, and it cannot be verified by the arbitrator or any other third party. Furthermore, the government cannot sell the right not to regulate to the investor.[10] Thus, the investor and the government cannot agree on any contract or agreement that has to rely on the realization of this stochastic variable, and without any arbitration, the potential relationship will be dominated by a serious holdup problem. Since the investor will leave if $\pi(a) < \bar{\pi}$, domestic welfare, denoted by V, is given by

$$V = \begin{cases} 0 & \text{if } \pi(a) < \bar{\pi}, \\ v(\theta, a) > 0 & \text{if } \pi(a) \geq \bar{\pi}. \end{cases} \tag{2}$$

We make the following

Assumption 1

$$\begin{aligned} & v(\theta, a) : v_a(\theta, 0) > 0, v_{aa}(\cdot) < 0, v_{\theta a}(\cdot) > 0, \exists \tilde{a}(\theta) > 0 : v_a(\theta, \tilde{a}(\theta)) = 0, \\ & v_a(\underline{\theta}, 0) + \pi_a(0) > 0, \\ & v_a(\overline{\theta}, \bar{a}) + \pi_a(\bar{a}) < 0, \\ \forall \theta \in [\underline{\theta}, \overline{\theta}] : & W(\theta, a^*(\theta)) = v(\theta, a^*(\theta)) + \pi(a^*(\theta)) - \Phi > 0 \\ & \text{where } a^*(\theta) = \arg\max_{a(\theta)} W(\theta, a(\theta)). \end{aligned}$$

W denotes aggregate welfare. Assumption 1 implies that domestic welfare is concave in a, that an increase in a increases the marginal welfare w.r.t. θ, and that a function $\tilde{a}(\theta)$ exists that maximizes domestic welfare for any realization of θ. Furthermore, some $a > 0$ is socially desirable, and it is never socially desirable to reduce investor profit to $\bar{\pi}$, while entry is always socially desirable.

Note that we do not need any assumption on $v_\theta(\cdot)$ for the subsequent analysis. A natural assumption would be that $v_\theta(\cdot)$ is negative such that a large realization of θ leads to a large direct drop in domestic welfare. What is more important, however, is that marginal welfare improves with θ such that the domestic government will want to fight a large θ-shock with a large a. As for θ, both the government and the investor anticipate that $\theta \in [\underline{\theta}, \overline{\theta}], 0 < \underline{\theta} < \overline{\theta} < \infty$, and they know that θ is distributed according to the c.d.f. $G(\theta)$. Expression (2) acknowledges that domestic welfare can be strictly positive only if the investor decides to stay in the country, and then domestic welfare

[9] Note that the model could also accommodate a change in political climate such that θ determines the government's objective function and thus measures political shocks.

[10] If it could, the holdup problem would disappear: the investor would hold the property rights to regulate and could sell them (efficiently) to the government, making ISDS redundant. However, we do not observe any such arrangement as each country could renege due to the sovereignty constraint.

is strictly positive even if $a = 0$. We find for the regulating activity that maximizes domestic welfare that

$$\tilde{a}'(\theta) = -\frac{v_{\theta a}(\theta, \tilde{a}(\theta))}{v_{aa}(\theta, \tilde{a}(\theta))} > 0,$$

holds, that is, regulating activities increase with the intervention necessity if domestic welfare is maximized.

The game has potentially five stages: In the first stage, the investor and the government bargain over an entry subsidy Σ paid to the investor (or an entry tax paid to the government if $\Sigma < 0$) and a potential ISDS agreement, followed by the second stage in which the firm decides on entry. Note that the first stage includes the case of no agreement which is equivalent to $\Sigma = 0$ and no ISDS provision. Without entry, the game is over, and with entry the intervention necessity is revealed to the host country and the investor in the third stage, after which the government announces its policy that can be challenged in case of an ISDS provision. In case of an ISDS provision, the ISDS protocol is applied if the government's policy has been challenged. Table 1 summarizes the structure of the game, and the paper will outline the details of an (optimal) ISDS mechanism in Sect. 3.

In any case, both the government and the investor will anticipate the potential future outcomes which allows us to scrutinize Stage I of the game. With Σ as the subsidy paid to the investor upon entry, the investor will enter if

$$\widehat{\pi} + \Sigma \geq \Phi, \widehat{\pi} = \int_{\underline{\theta}}^{\overline{\theta}} \pi(a(\theta)) dG(\theta),$$

holds where $\widehat{\pi}$ denotes the expected operating profit after entry. Both parties form rational expectations on $a(\theta)$ that will depend on whether an (optimal) ISDS provision is in place or not. The expected *ex ante* welfare of the host country is given by

Table 1 Game structure

Stage I:
The government and the investor have an agreement on an entry subsidy (tax) Σ with or without ISDS provision
Stage II:
The firm decides on entry. If it enters, it sinks the entry cost Φ
Stage III: The intervention necessity θ is revealed (only) to both parties
Stage IV:
Without ISDS provision, the host country chooses a, and the investor decides to stay or to leave. With an ISDS provision, the host country announces a and the firm decides to challenge a or not
Stage V:
If the firm challenges a, the ISDS protocol is applied. If the firm does not challenge a, a is applied

$$\widehat{V} = \widehat{v} - \Sigma, \widehat{v} = \int_{\underline{\theta}}^{\overline{\theta}} v(\theta, a(\theta)) dG(\theta).$$

If $\widehat{\pi} - \Phi + \widehat{v} < 0$, no welfare gains can be realized and no agreement will be reached. If $\widehat{\pi} - \Phi + \widehat{v} \geq 0$, an agreement can be reached, and we model stage I as a Nash bargaining game in which the government (the investor) has bargaining power of size η $(1 - \eta)$, $0 < \eta < 1$. Their outside options are zero as the investor makes zero profit without entry and the host country cannot benefit from the firm without entry. Maximization of the Nash bargaining product $\mathcal{N} = (\widehat{v} - \Sigma)^{\eta}(\widehat{\pi} + \Sigma - \Phi)^{1-\eta}$ w.r.t. Σ yields $\Sigma = (1 - \eta)\widehat{v} - \eta(\widehat{\pi} - \Phi)$, $\widehat{V} = \eta \widehat{W}$ and $\widehat{\pi} = (1 - \eta)\widehat{W}$ where $\widehat{W} = \widehat{\pi} - \Phi + \widehat{v}$ is the expected *ex ante* aggregate welfare. Not surprisingly, both parties have an interest in maximizing aggregate welfare.

What happens without any investor protection? Since the domestic government has no interest to make the investor leave the country if she had entered, it will maximize $v(\theta, a)$ with respect to the activity level a subject to $\pi(a) \geq \bar{\pi}$. If $\bar{a} < \tilde{a}(\theta)$, the government will set $a = \bar{a}$. In particular, if $\bar{\pi} = 0$, that is, if the investment is completely country-specific, the government *de facto* expropriates the foreign investor. If $\bar{a} \geq \tilde{a}(\theta)$, the government will set $a = \tilde{a}(\theta)$, so the investor will still realize a profit. Here, the holdup problem is in action as the domestic government cannot commit to a policy that will maximize aggregate welfare W instead of domestic welfare V only.

Welfare losses are of two different types. At the extensive margin, if $\widehat{\pi} - \Phi + \widehat{v} < 0$, no agreement will be reached and the firm will not enter although entry is socially desirable due to Assumption 1. At the intensive margin, if $\widehat{\pi} - \Phi + \widehat{v} \geq 0$, the firm will enter, but aggregate welfare is smaller than possible, and thus the *ex ante* expected welfare gains are smaller to begin with.

What would the investor and the government do if a third party could verify the intervention necessity θ and enforce a contract between the two parties? Since both parties have an interest in maximizing aggregate welfare and redistribute the expected aggregate welfare through a subsidy or tax in stage I, they would maximize W, and the first-best policy is determined by the first-order condition

$$W_a(\theta, a^*(\theta)) = v_a(\theta, a^*(\theta)) + \pi_a(a^*(\theta)) = 0 \tag{3}$$

which should hold for all $\theta \in [\underline{\theta}, \overline{\theta}]$. $a^*(\theta)$ would be agreed upon and the side-payments through subsidy Σ could distribute the *ex ante* aggregate welfare gain. Note that $a^*(\theta)$ is a function given the concavity of both the domestic welfare function and the profit function, and that $a^*(\theta) < \tilde{a}(\theta)$. Since $\pi_{aa}(\cdot) \leq 0$ and $v_{aa}(\cdot) < 0$, we also find that

$$a^{*\prime}(\theta) = -\frac{v_{\theta a}(\theta, a^*(\theta))}{v_{aa}(\theta, a^*(\theta)) + \pi_{aa}(a^*(\theta))} > 0 \tag{4}$$

holds, that is, that the optimal policy will lead to an increase in the regulating activity level with the intervention necessity, as to be expected. If the optimal policy could be implemented, the foreign investor would correctly anticipate that her expected profit

at this stage would be equal to $\int_{\underline{\theta}}^{\overline{\theta}} \pi(a^*(\theta))dG(\theta)$, and the (positive) entry decision would also be socially optimal at the investment stage, and the *ex ante* expected welfare gains would be maximal, too. In what follows, we investigate how the optimal policy (4) can be implemented using arbitration that cannot verify the intervention necessity.

3 Designing an optimal ISDS mechanism

We now develop an optimal ISDS mechanism. After a positive entry decision, the government will announce a which will be equivalent to announcing an intervention necessity, denoted by $\hat{\theta}$. This announcement cannot be proven true or false by any third party. However, the investor, being familiar with the implications of her investment, may challenge this announcement, and any challenge will have implications for both the investor and the government that will be managed by the arbitrator.

The proposed mechanism relies on a simple design and makes use of Myerson (1979) Revelation Principle: we can confine the analysis to type-dependent actions, and the design will be such that the country will (optimally) announce the true type of the intervention necessity. In particular, in stage 1, the government and the investor agree on the optimal action plan $a^*(\hat{\theta})$, on an alternative action plan $a^{**}(\hat{\theta})$ and transfer schemes F and $T(\hat{\theta})$. The alternative action plan fulfils: $\forall\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$: $a^*(\hat{\theta}) > a^{**}(\hat{\theta})$ and $a^{**\prime}(\hat{\theta}) > 0$. What is the purpose of this agreement? First, it should imply that the government has an incentive to truthfully reveal the intervention necessity. Second, it should prevent the investor to challenge the government when the government has announced the intervention necessity truthfully. We now develop such a mechanism to be managed by the arbitrator.[11] In detail, they agree on the following protocol potentially to be applied in stage V:

- *Step 1*: After the government has announced a^* which is equivalent to announcing the intervention necessity $\hat{\theta}$ due to $a^{*\prime}(\hat{\theta}) > 0$, the investor may challenge the announcement.
  - If she does not, action plan $a^*(\hat{\theta})$ is implemented, no transfers are paid and the game is over.
  - If she does, the government has to make an upfront payment of size F to the arbitrator.

[11] This is not the only mechanism that can do the job, but possibly most appealing as it imposes a minimum of requirements. An elaboration of more complex alternative mechanisms is available upon request.

- *Step 2*: If $\hat{\theta}$ has been challenged, the government is offered to continue with action plan $a^*(\hat{\theta})$ for which it will pay a transfer $T(\hat{\theta})$ to the arbitrator. Otherwise action plan $a^{**}(\hat{\theta})$ is implemented.
  - If the government accepts, action plan $a^*(\hat{\theta})$ is implemented and the government pays $T(\hat{\theta})$ to the arbitrator.
  - If the government rejects, action plan $a^{**}(\hat{\theta})$ is implemented and the arbitrator pays F to the investor.

Some comments are in order now. First, our restrictions on the choice of the alternative action plan are very mild. Second, note carefully that the arbitrator has at no point in this process to guess what the true realization of θ is. Third, since the arbitrator does not receive any signal, our mechanism has nothing to do with liability rules as they are known from trade dispute arbitration (see Beshkar, 2016). In our setup, the only qualification the arbitrator has to meet is that she has no own agenda, but follows the rules the investor and the government have agreed upon.

Why is it that the investor does not receive any compensation if the government wants to continue with the original optimal action plan? The reason is that we will design the transfer $T(\hat{\theta})$ such that the government will want to continue with $a^*(\hat{\theta})$ if and only if the investor should not have challenged the announcement. On the one hand, this is the reason why the investor does not receive any compensation as such a challenge should not make her better off. On the other hand, the investor does not have to be fined so no problem arises if the investor is financially constrained and cannot credibly commit to any payment.[12] In detail, suppose that the transfer in Stage 4 from the government to continue with $a^*(\hat{\theta})$ is given by

$$T(\hat{\theta}) = v(\hat{\theta}, a^*(\hat{\theta})) - v(\hat{\theta}, a^{**}(\hat{\theta})) > 0. \tag{5}$$

$T(\hat{\theta})$ depends only on the announcement, and since $v_a(\cdot) > 0$ and $a^*(\hat{\theta}) > a^{**}(\hat{\theta})$, $T(\hat{\theta})$ is unambiguously positive. Furthermore, we introduce

$$\Psi = \max_{\theta, \hat{\theta} \in [\underline{\theta}, \overline{\theta}]} \left\{ v(\theta, a^{**}(\hat{\theta})) - v(\theta, a^*(\theta)) \right\}.$$

This term determines the revelation condition an optimal ISDS mechanism will have to meet. It measures the difference between domestic welfare of the action plan after an over-reported intervention necessity has been successfully challenged and domestic welfare of truthfully announcing the intervention necessity. The sign of Ψ depends on the domestic welfare function and the design of the alternative action plan $a^{**}(\hat{\theta})$. It is likely to be positive, but we cannot rule out that $\Psi < 0$ if the

[12] This does not mean that a fine affecting the investor cannot be part of a more complex mechanism, for example due to some fairness consideration. See footnote 11.

difference in activity levels between the alternative action plan $a^{**}(\hat{\theta})$ and the optimal action plan $a^{*}(\hat{\theta})$ is very large.[13] We find:

Proposition 1 *If $T(\hat{\theta})$ is set according to (5) and $F \geq \Psi$ if $\Psi \geq 0$ or $F \geq 0$ if $\Psi < 0$, the ISDS mechanism will imply the optimal action plan $a^{*}(\theta)$ as a subgame-perfect equilibrium.*

Proof In order to qualify for an optimal mechanism, the mechanism has to meet several conditions. Since $v_a(\theta, a^{*}(\theta)) > 0$ and $a^{*\prime}(\cdot) > 0$, it is obvious that the government has no incentive to under-report θ: if $\hat{\theta} < \theta$, aggregate welfare would be lower than maximal, and it would only be the investor benefiting from this under-reporting while domestic welfare would decline. Hence, an announcement should be challenged by the investor only if the government over-reports θ. Following Aghion et al. (2018), we label these conditions appropriate challenge and inappropriate challenge, that is, a challenge by the investor is appropriate if $\hat{\theta} > \theta$ and is inappropriate otherwise. Suppose that the investor has challenged the announcement of the government. The transfer scheme $T(\hat{\theta})$ is designed such that $a^{**}(\hat{\theta})$ will be implemented only after an appropriate challenge. In particular, the government will accept to implement action plan $a^{**}(\hat{\theta})$ if only if

$$\begin{gathered} v(\theta, a^{**}(\hat{\theta})) > v(\theta, a^{*}(\hat{\theta})) - T(\hat{\theta}) \Leftrightarrow \\ v(\hat{\theta}, a^{*}(\hat{\theta})) - v(\hat{\theta}, a^{**}(\hat{\theta})) > v(\theta, a^{*}(\hat{\theta})) - v(\theta, a^{**}(\hat{\theta})) \Leftrightarrow \\ \hat{\theta} > \theta, \end{gathered} \tag{6}$$

where the last line follows from $v_{\theta a}(\cdot) > 0$.[14] Consequently, the offer to continue with plan $a^{*}(\theta)$ is rejected if and only if the government has been over-reporting θ. Thus, an inappropriate challenge can never imply the alternative action plan $a^{**}(\hat{\theta})$, and an appropriate challenge can never imply the optimal action plan $a^{*}(\hat{\theta})$. Furthermore, an appropriate challenge will be made if $\pi(a^{**}(\hat{\theta})) + F \geq \pi(a^{*}(\hat{\theta}))$ which is always fulfilled as $\pi(a^{**}(\hat{\theta})) > \pi(a^{*}(\hat{\theta}))$ and $F \geq 0$, and an inappropriate challenge will not be made because the investor cannot improve on her profit $\pi(a^{*}(\hat{\theta}))$ as the government will prefer to continue with $a^{*}(\theta)$. Furthermore, truthful revelation requires $v(\theta, a^{*}(\theta)) \geq v(\theta, a^{**}(\hat{\theta})) - F$ that is fulfilled due to the specification of F. □

Note carefully that the arbitrator will never run a deficit and could make a profit off the equilibrium. Furthermore, the mechanism could even include a legal cost to be carried by a challenging investor as long as the cost is not larger than

[13] If $v_a(\theta, a^{**}(\overline{\theta})) \geq 0$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$, $v(\theta, a^{**}(\hat{\theta}))$ will be largest for $\hat{\theta} = \overline{\theta}$. In this case, $\Psi < 0$ if $v(\theta, a^{**}(\overline{\theta})) < v(\theta, a^{*}(\theta))$ for all $\theta \in [\underline{\theta}, \overline{\theta}]$. This may happen if $a^{**}(\overline{\theta}) < a^{*}(\underline{\theta})$ such that the alternative action plan a^{**} specifies a lower activity level for the largest possible $\hat{\theta}$-announcement than the optimal activity plan a^{*} does for the smallest θ-realization, that is, for $\theta = \underline{\theta}$.

[14] In detail, let $\phi(x) = v(x, a^{*}(\hat{\theta})) - v(x, a^{**}(\hat{\theta}))$. Then, $\phi'(x) = v_x(x, a^{*}(\hat{\theta})) - v_x(x, a^{**}(\hat{\theta})) > 0$ because $v_{xa}(\cdot) > 0$ and $a^{*}(\hat{\theta}) > a^{**}(\hat{\theta})$. Thus, $\phi(x)$ increases with x and $\phi(x) > \phi(\theta)$ for all $x > \theta$.

$\pi(a^{**}(\hat{\theta})) + F - \pi(a^{*}(\hat{\theta}))$. In this case, the investor will make the challenge if it is appropriate and will not if it is not; in the latter case, the investor would even make a loss. The mechanism is also strategically equivalent to one in which the government has to pay a deposit of size F in the first stage that is paid back to the government in case of no challenge, stays with the arbitrator in case of an inappropriate challenge and is paid to the investor in case of an appropriate challenge.[15]

Another interesting feature of this simple mechanism is that the two parties have a choice how to arrange investment protection. If they choose an alternative action plan that implies a substantial reduction in a, the fine F will be small, but the transfer T has to be large in order to rule out that the government can continue with a^* too easily. If they choose an a^{**} close to a^*, the fine F will be substantial as to guarantee a true θ announcement, but the transfer T will be small. If the government's liability is limited, it may imply that both F and T are constrained as well. In this case, the ISDS mechanism is still feasible if F and T exist such that they do not violate the liability constraint.

In any case, the notion of an ISDS compensation in this optimal design does not match the ISDS provisions as they are employed in bilateral investment treaties or had been suggested for multilateral agreements like TTIP and TPP. For example, TPP was designed to indemnify an investor for "unjustified" profit losses. In an optimal design setup, an investor will not receive any compensation in equilibrium. But even if the government did over-report the intervention necessity, the compensation the investor would receive would be equal to F, but F is determined by domestic welfare effects only. Consequently, any design that wants the government to truthfully report the intervention necessity must rely on domestic welfare effects only.

What happens if the investor cannot exactly observe the true intervention necessity? Suppose that the investor receives a signal $\tilde{\theta}$ drawn from a distribution with support $[\theta - \epsilon, \theta + \epsilon]$ where θ is the true intervention necessity and $\mathrm{E}(\tilde{\theta}) = \theta$. If ϵ is not too large, the government and the investor could agree on schedules $a^*(\theta_i)$ and $a^{**}(\theta_i)$ where $i \in \mathbb{N}$ and $\theta_i > \theta_{i-1} + 2\epsilon$. This means that the agreement is confined to a limited number of intervention necessities such that not functions but simple schedules will be agreed. If the signal received by the investor is not too noisy, that is, if ϵ is not too large, this schedule will produce a countable list from the functions $a^*(\theta)$ and $a^{**}(\theta)$ from above, and the welfare losses would be rather small compared to the approach above.

4 An example

The general model has developed a mechanism that is immune against any defection from true announcement, but has not scrutinized the optimal defection options. In order to do so, this section considers an example for which domestic welfare is linear-quadratic and foreign profit is linear in a. In particular,

[15] See Gerber and Wichardt (2009) for a model in which players pay a deposit before making a contribution to a public good.

$$v(\theta, a) = \theta a - a^2/2 + \delta - \gamma\theta, \pi(a) = \pi_0 - \beta a, \bar{\pi} = 0, \beta < \underline{\theta}, \delta > \gamma\overline{\theta}, \pi_0 > \beta\overline{\theta}, \gamma > 0.$$

A straightforward implication of this specification is that the non-cooperative activity level is given by $\tilde{a}(\theta) = \theta$, as the foreign profit is always positive even without investor protection, and the investment has no value outside the host country, that is, $\bar{\pi} = 0$. The globally optimal activity level is determined by $a^*(\theta) = \theta - \beta$. Note that domestic welfare has also a fixed component δ that measures the fixed benefit of a foreign investor being active in the domestic country. Additionally, a high θ-realization reduces this benefit directly, but $\delta > \gamma\overline{\theta}$ guarantees that the government has never an interest to see the investor leave the country.

Furthermore, suppose that the government and the investor agree upon a simple reduction method to determine the difference between $a^{**}(\hat{\theta})$ and $a^*(\hat{\theta})$:

$$a^{**}(\hat{\theta}) = a^*(\hat{\theta}) - \alpha, a^*(\underline{\theta}) > \alpha > 0, \overline{\theta} - \underline{\theta} > \beta + \alpha. \tag{7}$$

This is a very simple agreement as an appropriate investor challenge will imply a reduction of a by α. Furthermore, (7) assumes that the spread of the θ-realizations is sufficiently large; this assumption will guarantee that Ψ will be an interior maximum. The agreement (7) implies that the transfer according to (5) does not depend on $\hat{\theta}$ and is equal to

$$T(\hat{\theta}) = \bar{T} = \frac{\alpha(2\beta + \alpha)}{2}.$$

This is the transfer that the government has to pay for continuing with the optimal action plan $a^*(\hat{\theta})$. Note that $\bar{T}$ can be made very small by choosing a small α. The true revelation condition depends on both the true intervention necessity θ and the announced intervention necessity $\hat{\theta}$. Our specification allows us to scrutinize the optimal defection options of the government. Since transfers do not depend on $\hat{\theta}$ in our example, the government will do best if it chooses $\hat{\theta}$ as to maximize $v(\theta, a^{**}(\hat{\theta}))$ with respect to $\hat{\theta}$ subject to $\hat{\theta} \in [\underline{\theta}, \overline{\theta}]$, given that it decides to over-report. Differentiation of $v(\theta, a^{**}(\hat{\theta}) = \theta(\hat{\theta} - \beta - \alpha) - (\hat{\theta} - \beta - \alpha)^2/2$ shows that the optimal announced (over-reporting) intervention necessity is given by

$$\hat{\theta}^*(\theta) = \min\{\theta + \beta + \alpha, \overline{\theta}\}.$$

The optimal defection will always aim at maximizing domestic welfare only, and it can undo the positive effect for the foreign investor and the effect of the alternative action plan by claiming that θ is larger by $\beta + \alpha$ than it actually is. Since the government cannot announce any $\hat{\theta}$ that is out of range, this over-reporting works only for sufficiently small θ realizations, but is constrained for large θ realization by $\hat{\theta} \leq \overline{\theta}$ if $\theta + \beta + \alpha > \overline{\theta}$. We find that

$$v(\theta, a^{**}(\hat{\theta}^*(\theta))) - v(\theta, a^*(\theta)) = \begin{cases} \beta^2/2 & \text{if } \theta \leq \overline{\theta} - \beta - \alpha, \\ \dfrac{\beta^2 - \theta^2 + (\overline{\theta} - \beta - \alpha)^2}{2} & \text{if } \theta > \overline{\theta} - \beta - \alpha. \end{cases}$$

Note that $v(\theta, a^{**}(\hat{\theta}^*(\theta))) - v(\theta, a^*(\theta))$ is smaller for $\theta > \overline{\theta} - \beta - \alpha$ than for $\theta \leq \overline{\theta} - \beta - \alpha$,[16] and thus $\Psi = \beta^2/2$; so Ψ is independent of α. Consequently, the true revelation condition is given by

$$F \geq \Psi = \beta^2/2,$$

and our example shows that very simple rules can be applied if domestic welfare can be approximated by a linear-quadratic function and maximized profits can be approximated by a linear function.

5 Concluding remarks

This paper has developed a simple optimal ISDS mechanism. This mechanism goes beyond the general provisions as found in bilateral and multilateral investment treaties. First, any ISDS compensation should be based on the host country's welfare effects, and not on any foregone profit. The paper does in no way indicate that the suggested ISDS mechanism is politically feasible. On the one hand, it requires a thorough cost-benefit analysis within host countries, commitment to a transfer scheme and action plans, and it relies on arbitration. However, recent designs that employ ISDS tribunals also require commitment to these ISDS rules, so any investor protection will be infeasible without a minimum degree of commitment.

On the other hand, detailed and transparent arbitration rules may even make any investor protection provision more agreeable by the public. It may be held against the optimal mechanism that the investor must be able to learn the intervention necessity. But if she cannot (and therefore the proposed mechanism is not feasible), it is hard to see how a tribunal can do a better job. It also seems to be a fair assumption that investors know the impact of their activities not only on their profits, but also on other agents. Furthermore, we could also show that the mechanism is also implementable if the investor receives a noisy signal as long as the noise is not too large.

The main limitation of this approach is the requirement to be able to foresee all possible outcomes of all possible investments. If this is not possible and/or if the outcomes are too complex to be codified, designing optimal ISDS provisions for all investment activities is not feasible. But given this backdrop, it does not mean that some or even many investment outcomes can be specified such that a general ISDS provision can be complemented and even substituted by more detailed provisions.

Furthermore, an efficient investment protection provision must involve three parties and needs to be managed credibly by an arbitrator. The paper has built upon

[16] $\beta^2/2$ does not change with θ, but differentiating $(\beta^2 - \theta^2 + (\overline{\theta} - \beta - \alpha)^2)/2$ with respect to θ yields $-\theta < 0$. $\overline{\theta} - \beta - \alpha > \underline{\theta}$ due to (7).

the implementation literature to design the optimal mechanism. However, as discussed by Maskin (2002), subgame-perfect implementation mechanisms are hardly ever used to tackle the holdup problem in private markets, so a market in which arbitrators offer their services to solve the holdup problem efficiently does obviously not exist. He concludes that there must be other frictions that play a role for the non-existence of these markets. Additionally, the possibility of renegotiations between two parties is known to jeopardize an *ex ante* efficient agreement if the two parties can do better without arbitrator *ex post*.

In our context, however, well-established supranational institutions like the International Centre for Settlement of Investment Disputes and the International Centre for Settlement of Investment Disputes already serve as an arbitrator in ISDS disputes. These institutions could also manage more detailed provisions in the sense of this paper and make renegotiations difficult. Moreover, the WTO is already an expert in arbitration through its Dispute Settlement Body (DSB) that deals with trade disputes. It thus seems that supranational institutions must be given a more active role to make investor protection successful and acceptable. In conclusion, there is no reason why agreements should not be able to offer an institutional platform that a host government and a foreign investor can use to agree on an mechanism like the one suggested above.

Acknowledgements This paper has been presented at several seminars, and I am grateful for comments and suggestions, in particular to two anonymous referees, Ngo Van Long and Guillaume Lord. I also gratefully acknowledge financial support received from the Australian Research Council under Project Number DP190103524. I dedicate this paper to my dear friend and colleague Ngo Van Long who passed away on January 15, 2022.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

References

Aghion P, Fehr E, Holden R, Wilkening T (2018) The role of bounded rationality and imperfect information in subgame perfect implementation—an empirical investigation. J Eur Econ Assoc 16:232–274

Aisbett E, Karp L, McAusland C (2010a) Compensation for indirect expropriation in international investment agreements: implications for national treatment ad rights to invest. J Glob Dev 1(2), Article 6

Aisbett E, Karp L, McAusland C (2010b) Police powers, regulatory takings and the efficient compensation of domestic and foreign investors. Econ Record 86:367–383

Aisbett E, Bonnitcha J (2021) A pareto-improving compensation rule for investment treaties. J Int Econ Law 24:181–202
Antràs P (2003) Firms, contracts, and trade structure. Q J Econ 118:1375–1418
Antràs P, Helpman E (2004) Global sourcing. J Polit Econ 112:552–80
Antràs P, Chor D (2013) Organizing the global value chain. Econometrica 81:2127–2204
Bagwell K, Staiger RW (2005) Multilateral trade negotiations, bilateral opportunism and the rules of GATT/WTO. J Int Econ 67:268–294
Bagwell K, Bown CP, Staiger RW (2016) Is the WTO passé? J Econ Lit 54:1125–1231
Berger A, Busse M, Nunnenkamp P, Roy M (2011) More stringent BITs, less ambiguous effects on FDI? Not a bit! Econ Lett 112:270–272
Beshkar M (2010) Optimal remedies in international trade agreements. Eur Econ Rev 54:455–466
Beshkar M (2010) Trade skirmishes and safeguards: a theory of the WTO dispute settlement process. J Int Econ 82:35–48
Beshkar M (2016) Arbitration and renegotiation in trade agreements. J Law Econ Organ 32:586–619
Blume L, Rubinfeld D, Shapiro P (1984) The taking of land: when should compensations be paid? Q J Econ 99:71–92
Bonnitcha J, Aisbett E (2021) Against balancing: revisiting the use/regulation distinction to reform liability and compensation under investment treaties. Mich J Int Law 42:231–290
Busse M, Königer J, Nunnenkamp P (2010) FDI promotion through bilateral investment treaties: more than a bit? Rev World Econ 146:147–177
Egger P, Merlo V (2012) BITs bite: an anatomy of the impact of bilateral investment treaties on multinational firms. Scand J Econ 114:1240–1266
Egger P, Pfaffermayr M (2004) The impact of bilateral investment treaties on foreign direct investment. J Comp Econ 32:788–804
Gerber A, Wichardt PC (2009) Providing public goods in the absence of strong institutions. J Public Econ 93:429–439
Haftel Y (2010) Ratification counts: US investment treaties and FDI flows into developing countries. Rev Int Polit Econ 17:348–377
Hermalin B (1995) An economic analysis of takings. J Law Econ Organ 11:64–86
Horn H, Tangerås T (2016) Economics and politics of international investment agreements, IFN Working Paper No. 1140, Research Institute of Industrial Economics
Janeba E (2019) Regulatory chill and the effect of investor state dispute settlements. Rev Int Econ 27:1172–1198
Klimenko M, Ramey G, Watson J (2007) Recurrent trade agreements and the value of external enforcement. J Int Econ 74:475–499
Kohler W, Stähler F (2019) The economics of investor protection: ISDS versus National Treatment. J Int Econ 121, article 103254
Konrad K (2017) Large investors, regulatory taking and investor-state dispute settlement. Eur Econ Rev 98:341–353
Maggi G, Staiger R (2011) The role of dispute settlement procedures in international trade agreements. Q J Econ 16:475–515
Maggi G, Staiger R (2015) Optimal design of trade agreements in the presence of renegotiation. Am Econ J Microecon 7:109–143
Maggi G, Staiger R (2018) Trade disputes and settlement. Int Econ Rev 59:19–50
Maskin E (2002) On indescribable contingencies and incomplete contracts. Eur Econ Rev 46:725–733
Maskin E, Tirole J (1999) Unforeseen contingencies and incomplete contracts. Rev Econ Stud 66:83–114
Miceli TJ, Segerson K (1994) Regulatory takings: when should compensation be paid? J Leg Stud 23:749–776
Moore J, Repullo R (1988) Subgame perfect implementation. Econometrica 56:1191–1220
Myerson R (1979) Incentive-compatibility and the bargaining problem. Econometrica 47:61–73
Navaretti GB, Venables AJ (2006) Multinational firms in the world economy. Princeton University Press, Princeton
Neumayer E, Spess L (2005) Do bilateral investment treaties increase foreign direct investment to developing countries? World Dev 33:1567–1585
Nosal E (2001) The taking of land: market value compensation should be paid. J Public Econ 82:431–443
OECD (2012) Investor-state dispute settlement, public consultation: 16 May–9 July 2012, Paris
Park J (2011) Enforcing international trade agreements with imperfect private monitoring. Rev Econ Stud 78:1102–1134

Rose-Ackerman S, Tobin J (2011) When BITs have some bite: the political-economic environment for bilateral investment treaties. Rev Int Organ 6:1–32

Rose-Ackerman S, Tobin J (2009) Do BITs benefit developing countries? In: Rogers CA, Alford RP (eds) The future of investment arbitration. Oxford University Press, Oxford

Schjelderup G, Stähler F (2021) Investor state dispute settlement and multinational firm behavior. Rev Int Econ 29:1013–1024

TPP draft (2016) Text of the trans-pacific partnership, New Zealand Ministry of Foreign Affairs and Trade. https://www.mfat.govt.nz/en/about-us/who-we-are/treaties/trans-pacific-partnership-agreement-tpp/text-of-the-trans-pacific-partnership/. Accessed 26 Feb 2022