RESEARCH INSTITUTE OF INDUSTRIAL ECONOMICS

IFN Working Paper No. 1140, 2016

Economics and Politics of International Investment Agreements

Henrik Horn and Thomas Tangerås

Research Institute of Industrial Economics
P.O. Box 55665
SE-102 15 Stockholm, Sweden
info@ifn.se
www.ifn.se

Economics and Politics of International Investment Agreements[1]

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

Thomas Tangerås
The Research Institute of Industrial Economics (IFN), Stockholm

November 18, 2016

[1] The authors gratefully acknowledge financial support from the Marianne and Marcus Wallenberg Foundation (Horn), Riksbankens Jubileumsfond (Horn) and Jan Wallanders och Tom Hedelius Stiftelse (Tangerås).

Abstract

This paper investigates the design and implications of international investment agreements. These are ubiquitous, potent and heavily criticized state-to-state treaties that protect foreign investment against host country policies. We show that optimal agreements cause national but not global underregulation ("regulatory chill"). The incentives to form agreements and their distributional consequences depend on countries' unilateral commitment possibilities and the direction of investment flows. The benefits from agreements between developed countries accrue to foreign investors at the expense of the rest of society, but this is not the case for agreements between developed and developing countries.

JEL Codes: F21; F23; F53; K33
Keywords: Foreign direct investment; expropriation; international investment agreements; regulatory chill

Contents

1 Introduction

International investment agreements are state-to-state treaties that aim to promote foreign direct investment (FDI) by protecting investors against adverse host country policy measures. The agreements typically require host countries to compensate foreign investors for expropriation and measures with similar effects, and they contain a range of additional contract provisions, such as non-discrimination of foreign investment. They often include dispute settlement mechanisms that enable foreign investors to litigate against host countries through legal processes outside the domestic legal system, so called investor-state dispute settlement (ISDS).

The first investment agreements appeared in late 1950s, but most of the 2 600 treaties that currently are in force were formed after 1990.[1] A majority of the agreements exclusively address investment protection, but it has become increasingly common also for preferential trade agreements to encompass such protection. The North American Free Trade Agreement (NAFTA) was one of the first trade agreements to do so, and this has become a standard feature of EU and US preferential trade agreements.

The agreements have until recently been formed without much political opposition, but investment protection has recently come under intense fire. The debate concerns in particular the role of investment protection in the "mega-regional" preferential trade agreements currently under negotiation or in the process of ratification—the Trans-Pacific Partnership (TPP), the Canada-EU Comprehensive Economic and Trade Agreement (CETA), and the EU-US Transatlantic Trade and Investment Partnership (TTIP). Opponents argue in particular that the investment protection chapters included in these agreements impose undemocratic constraints on the ability of signatory states to pursue legitimate policy goals—that is, that they cause "regulatory chill."

The public distrust of investment protection has been fueled by a number of litigations that have made headlines: TransCanada Corporation recently declared its intention to sue the US under NAFTA for US$ 15 billion as compensation for the US decision to disallow the construction of the Keystone XL pipe line. Phillip Morris litigated against Australia over the tobacco plain packaging legislation (but lost). Spain is facing a large number of litigations for the withdrawal of renewable energy support schemes during the financial crisis of 2008, and similar cases have been brought against Italy and the Czech Republic. Germany is being sued by the energy company Vattenfall for losses caused by the country's decision to shut down nuclear power in the wake of the Fukushima disaster. Many see these cases as clear indications of a flawed investment protection regime.

At heart of the skepticism lies the fact that agreements protect investors not only against direct expropriation, but also indirect (or regulatory) expropriation. Indirect expropriations can arise when host countries take regulatory measures that significantly reduce the value of the investment for its owners, but do not entail a change in the ownership of the assets.[2] Critics argue that

[1] http://investmentpolicyhub.unctad.org/IIA.

[2] Another criticized type of substantive obligation is the "fair and equitable treatment" provision that is included in most agreements.

the rules concerning indirect expropriation are so vaguely formulated that almost any regulatory policy with adverse consequences for foreign investors could be interpreted to constitute an indirect expropriation. This is seen to be particularly troublesome given the possibility for investors to use the very potent ISDS mechanisms to enforce the substantive obligations in the agreements.[3]

The policy debate raises a number of questions concerning investment agreements: How should the agreements be designed? For instance, when should investors get compensation for losses that stem from regulatory interventions, and when should these be considered as normal business risk? Will appropriately designed agreements cause regulatory chill? When will they be formed? Who benefits and who loses from the formation of the agreements? The economic literature sheds very little light on these issues. The purpose of this paper is to contribute to filling this lacuna.

Our paper focuses mainly on regulatory expropriations. Central to the problem of optimally designing investment protection against regulatory expropriations is the interaction between two distortions: the host country disregard of foreign investor interests when deciding on regulation, and the investment behavior that creates a motive for regulation in certain instances. But the interaction between these distortions can cause overregulation and underinvestment, and thus create a scope for an investment agreement. To capture this interaction, we consider the negotiation of an investment agreement between two countries within a generalized version of the canonical regulatory takings model of investment protection.[4] We will assume that investment agreements are Pareto efficient, and that they are only entered into if they strictly benefit the countries to the agreement. Since we are ultimately interested in the implications of real world investment agreements, we constrain the agreements under study to share certain basic features with actual agreements. This is in our view a natural starting point for the analysis of investment agreements, but as explained below, the small related literature on investment agreements have followed different approaches in this regard.

At the outset of the interaction, competitive firms make irreversible foreign direct investment in production facilities. The investments create benefits to the host country, but also give rise to externalities of unknown magnitude at the time of the investment. These externalities may render production ex post undesirable from a national and perhaps even an international perspective. These shocks could represent a broad range of exogenous events, for instance the arrival of information concerning adverse environmental or heath consequences of the production process; we will simply denote this realization as a "regulatory shock." Upon observing the country-specific shock, each host country decides whether to permit or to disallow production. Production and consumption occurs if there is no regulation. The private investment is effectively lost if the country instead regulates the industry, even though there is no formal take-over of the ownership of the assets.

An investment agreement in this setting represents a set of negotiated rules specifying payments to foreign investors as a function of regulatory decisions, and possibly other factors. Importantly,

[3]Several other aspects of the adjudication system are also severely criticized, such as the lack of independence of arbitrators, the lack of appeal possibilities and excessive confidentiality.

[4]The "regulatory takings" concept stems from the Fifth Amendment to the US Constitution stating "...nor shall private property be taken for public use, without just compensation." See Section 1.1 for a discussion of the literature.

we require the set of potential compensation schemes to be congruent with a number of features of actual agreements. For instance, firms are only eligible for compensation in case of regulation, and any compensation must equal firms' foregone operating profits. These restrictions on the set of feasible agreements are intended to make model predictions more plausible as depictions of actuality. But we also identify circumstances under which the restrictions do not constrain agreements in terms of global efficiency.

Our first finding addresses a key concern in the policy debate, namely that investment agreements cause regulatory chill. We show that Pareto optimal investment agreements never yield underregulation from a joint welfare perspective—there will be no "global regulatory chill." When compensation is limited to at most each firm's operating profit, it is always optimal for the host country to regulate if it is globally optimal to do so. Instead, agreements induce either global overregulation or ex post optimal regulation. They also yield less regulation than would result without any agreement. Such "national regulatory chill" is simply the price the host country has to pay to promote foreign investment. These results hold for a large variety of settings and are fundamental to Pareto optimal investment agreements. For instance, they do not depend on market structure, nor do we have to restrict compensation to be equal to operating profit.

We next show that a very simple scheme whereby firms receive full compensation for all regulatory shocks below and no compensation above a specified threshold, is sufficient to implement any Pareto optimal compensation scheme when compensation is proportional to operating profit. This is known as a *carve-out policy*. We subsequently refer to this threshold as the level of investment protection. This finding simplifies the subsequent analysis of the formation and the implications of investment agreements considerably, since the agreements can be characterized in terms of the level of investment protection they offer. Generally speaking, the optimal level of investment protection trades off the marginal benefit to investors of increased protection against the marginal costs to the host country of excessive investment and underregulation.

We next turn to the formation of Pareto optimal investment agreements and their implications. The net benefit to a host country of an agreement depends on two fundamental factors. The first is whether the agreement yields improved commitment possibilities in terms of enforcing investment protection. It seems reasonable to assume that developed countries by and large are able to implement unilaterally credible investment protection schemes through laws and regulations, whereas developing countries typically do not have this capability. The second fundamental factor is whether the agreement allows for the internalization of international externalities from unilateral regulatory decisions. Again, there appears to be a clear distinction between developed and developing countries, in that developed economies are both host and source countries for FDI, whereas developing countries mainly are recipients of investments. We consider two archetypical settings that appear particularly policy relevant and help us distil the gains from investment agreements.

A North-South agreement is a treaty between a developed country (North) and developing one (South). South is not able to make a credible unilateral commitment to investment protection, and the purpose of an agreement is to stimulate investment from North to South.[5] If such an agreement is formed, it will by necessity increase domestic welfare in South—that is, the joint welfare generated in, and accruing to South— since this is the only reason for South to participate. Those benefits then stem from the external legal enforcement of investment protection commitments that the agreement offers: South would have nothing to gain from an agreement if it already had full commitment possibilities because then its unilaterally chosen level of investment protection would internalize all relevant effects of FDI. This mechanism corresponds closely to the "commitment approach" to trade agreements, which sees trade agreements as a tool to help governments withstand domestic protectionist pressures.

The other archetypical treaty we consider is a North-North agreement. Such an agreement adds little in terms of improved commitment possibilities. Instead, the gains stem from a bilateral internalization of the externalities from regulation. Each country achieves improved protection of its outgoing investment by offering foreign-owned industries better protection at home. This mechanism is much in line with the standard view of trade agreements, according to which they resolve Prisoners' Dilemmas by allowing countries to exchange increased exports against imports at mutual benefit. Contrary to the case of the North-South agreement, such improved investment protection uniquely benefits investors because the North-North agreement entails protection levels that are excessive from a domestic viewpoint and therefore reduce domestic welfare in both countries.

We believe these findings are informative about the politics of investment protection. The results suggest that symmetric agreements such as CETA and TTIP (and to some extent also the TPP) benefit foreign investors, but reduce consumer welfare in a broad sense of the term. This explains why the industry is in favor, but we witness such popular resistance to the formation of these agreements. Our findings also predict that there should be less opposition to North-South agreements, since the benefits to a larger extent accrue to the broader public in the host country. Again, this seems compatible with what we observe. Of course, there has been opposition also to North-South agreements, and some agreements have been renegotiated or revoked.[6] But it appears as if there for the most part has been less discontent with investment agreements in developing countries.

The remainder of the paper extends the analysis in a number of directions. First, to examine the common allegation that investment agreements undermine the sovereignty of democratically elected governments, we assume that the total shock is the product of an exogenous regulatory shock, such as a scientific discovery, and shocks to political preferences concerning the regulatory objective. Government preferences are unknown at the investment stage, but are realized simultaneously with the scientific shock. We show how welfare optimal compensation schemes are "democratic" in the

[5] The vast majority of bilateral investment agreements are between a developed and a developing country. For instance, the US has approximately 60 investment treaties with low and middle income developing countries; see the above-mentioned UNCTAD website.

[6] Stiglitz (2008) summarizes much of this critique from a developing country perspective, and provides a number of proposals for the design if IIAs.

sense of allowing governments that are more sensitive to regulatory shocks to intervene for a larger range of shocks without paying compensation.

Second, investment agreements typically include National Treatment (NT) provisions that prohibit a less favorable treatment of foreign investment than that of domestic investment undertaken in "like circumstances." We show that host governments can benefit from including NT clauses in investment agreements as a commitment tool to enforce stricter regulation of domestic industries.

Third, investment agreements typically have stricter rules for direct expropriation, in that they often do not include the general exceptions that apply to regulation. Allowing uncompensated direct expropriation might actually enhance efficiency. Direct expropriation separates the problem of correcting investment incentives from that of ensuring ex post optimal regulation incentives. Direct expropriation and regulation have the same consequences for the targeted firms, but the government will allow production in a seized asset if and only if doing so is welfare optimal. However, this mechanism depends crucially upon the regulatory shock being observable. Under asymmetric information, a host country government would always claim that the realization of the shock was such that it could seize the assets without compensation. Full compensation for all direct expropriation is the easiest way to avoid that investment is driven completely out of the market in this case.

Most of the analysis is conducted assuming that firms neglect all price effects and treat the probability of regulation as exogenous to their own investment. In a fourth extension we show that the optimality of a compensation scheme based on a threshold investment protection level is independent of the market structure.

A fifth extension assumes that the regulatory shock is no longer ex post verifiable, but instead private information to the host country. A mechanism with zero compensation for large shocks is incentive incompatible under such asymmetric information because the host country could then intervene at no domestic cost by exaggerating the severity of the regulatory shock. Incentive compatibility instead requires that the country pays a fixed compensation for all regulation. This compensation generally depends on the cost to the host country of maintaining production, a step away from the contract stipulations usually found in investment agreements.

As a final extension we allow for more general compensation schemes than those commonly included in investment agreements, in particular for the purpose of exploring the consequences of asymmetric information. Incentive compatible compensation schemes usually imply excess compensation by the host country (punitive damages) or third-party participation; see our below review of the literature. We show how to implement the first-best outcome by means of a relative performance scheme that neither involves punitive damages nor third-party payments.

The general picture that emerges from our analysis is that even when the treaty design is restricted to share basic features of actual agreements, there is still often scope for investment agreements to improve national welfare of the partner countries. But investment protection schemes in agreements such as CETA, TPP, and TTIP, are likely to have significant distributional consequences, plausibly being an important reason for the wide-spread political resistance to such agreements.

1.1 Relation to the literature

The economic literature on the incentives for expropriation of foreign investment dates back at least to Keynes (1924), but remained informal until the early 1980s.[7] Eaton and Gersovitz (1983, 1984) are among the first to study expropriation in a conventional neoclassical framework. Other authors highlight strategic aspects of the host country-investor relationship, focusing on implicit mechanisms rather than international treaties for reducing hold-up problems. For instance, Dixit (1988) informally sketches an incomplete information model of the interaction between a sequence of potential investors and a host country, in a situation where the host country preferences regarding expropriate are unknown to the investors (Raff (1992) formally analyzes such a setting). Dixit (1988) explains why host countries with a preference for expropriation might abstain in order to persuade investors that it is reputable, and why host countries will expropriate only toward the end of a finitely repeated game. Cole and English (1991) show how the incentives for expropriation can be kept at bay by the use of trigger strategies in an infinite horizon model. Thomas and Worral (1994) and Schnitzer (1999) examine how other forms of self-enforcing agreements between investors and host governments can remedy hold-up problems. Dixit (2011) discusses a range of issues related to insecurity of property rights and FDI, and also provides extensive reviews of both the theoretical and the empirical literature.

A set of more recent papers specifically addresses investment agreements as a solution to investor-state hold-up problems, but this literature is still small and fragmented. Markusen (1998, 2001) discusses pros and cons of investment agreements from a developing country perspective. Turrini and Urban (2000, 2008) analyze the role of a multilateral investment agreement. Another strand of literature examines whether investment agreements actually promote investment; see Lejor and Salfi (2015) for a recent contribution and a critical survey of this particular literature.

The first paper to examine the optimal design of investment agreements is Aisbett et al (2010a) who generalize a regulatory takings model to incorporate an imperfectly unobservable regulatory shock. Key contributions in the takings literature implicitly assume that the incentives to invest and to regulate are undistorted.[8] Aisbett et al (2010a) show how to achieve full efficiency when the incentives to regulate are distorted, if the host country can overcompensate the industry for its losses. Aisbett et al (2010b) examine a setting where the host country requests foreign firms to make up-front payments before they invest, and where a National Treatment clause renders such payments are infeasible. The presence of an NT rule can make it desirable to increase the

[7] For instance, Bronfenbrenner (1955) argues that expropriations are economically advantageous for many developing countries, and that developed country investors therefore should refrain from investing in such countries. A hold-up problem is at the core of Vernon's (1971) "obsolescing bargaining" theory of expropriation.

[8] See Blume et al (1984) and Miceli and Segerson (1994). In such instances, a compensation mechanism can only reduce welfare. Hermalin (1995) demonstrates how taxes and other sophisticated compensation mechanisms can achieve the first-best outcome in a takings model with distorted investment and regulation decisions. See Miceli and Segerson (2011) for a comprehensive survey. We discuss some of these papers in more detail in Section 5.6.

carve-out from compensation requirements for foreign firms. Stähler (2016) derives a mechanism that can achieve the first-best outcome when the regulatory shock is unobservable. His proposed solution breaks the payment balance between the host country and firms. The present paper differs from these contributions in several regards. For instance, all three studies rely on compensation schemes that are not found in actual investment agreements: subsidization and overcompensation are explicitly ruled out in important treaties (for instance TPP), and the agreements do not give any scope for breaking the budget balance by payments to or from third parties. Furthermore, none of the three papers discuss distributional effects of compensation schemes or the incentives to introduce them. Considering distributional effects is important to be able to understand the formation of investment agreements. For instance, a compensation mechanism that achieves the global optimum need not be an equilibrium outcome if it for the most part benefits foreign investors at the loss of the host country.

There is also an emerging literature that considers the implications of exogenously imposed investment agreements, and that sheds complementary light on some of the issues analyzed here. Bergstrand and Egger (2013) depict a investment agreement as an exogenous reduction in the capital cost of FDI in a three-factor, three country general equilibrium setting. The purpose is to analyze how the welfare gains of investment agreements and preferential trading agreements depend on factors such as country size and trade costs.

Janeba (2016) focuses on two sources of deficiencies in the arbitration process under investment agreements—litigation costs and biased courts. First, litigation costs can dissuade host countries from pursuing efficient policies. In such instances there is a form of regulatory chill. Second, countries might lose from unilateral shift to a system of international arbitration if arbitration courts appointed under investment agreements are more likely than domestic courts to rule in favor of foreign investors. But both countries can still benefit from such a system, provided that investment flows are sufficiently symmetric, since each country will receive more favorable treatment of the own investment abroad. We assume away implementation problems by considering full commitment to the negotiated agreement.

Kohler and Stähler (2016) examine consequences of a particular interpretation of the term "legitimate expectations," which occasionally has been adopted by arbitration panels. It holds that current regulatory policies can create legitimate investor expectations about future regulations and thus effectively link regulatory decisions across time. The authors show in a two-period framework how such intertemporal linkages can reduce overregulation and increase aggregate welfare over time. This type of agreement, denoted "ISDS," is then compared with an agreement that instead imposes a National Treatment rule that equalizes the protection of foreign and domestic firms. The authors identify circumstances under which this non-discrimination rule yields higher aggregate welfare than their "ISDS" mechanism. The present paper instead considers the endogenous interaction of NT clauses and negotiated levels of investment protection.

Schjelderup and Stähler (2016) investigate a two-period regulation problem in which a host country taxes a foreign investor to reduce a negative investment externality and raise tax revenue. The second period externalities are unknown and might require the host country to increase the tax. An "ISDS" mechanism requires the host country to set its taxes at a Pigouvian level and might request the country to compensate the firm for tax increases. The authors show that the mechanism could cause overinvestment and has potentially ambiguous welfare implications.

Konrad (2016) considers the strategic incentives for foreign firms to invest in order to reduce the probability of environmental regulation. Increased investment protection benefits domestic and foreign investors, but has an overall negative effect on the host country by exacerbating an already existing overinvestment and underregulation problem. Konrad (2016) sees these results as one explanation for why industries favor investment protection and those mechanisms are disliked by environmentalists and other interest groups. We derive our main results in a setting that allows for such considerations, but the results do not depend on strategic investor behavior.

The welfare analyses in the above papers rest in each case upon a comparison of an investment agreement with exogenously given characteristics to some outside option. There is no analysis of whether an alternative design of the agreement would make it welfare enhancing and acceptable to the parties if the proposed agreement is not, nor is there any discussion about the relevant alternative. Conversely, even if the studied agreement would increase aggregate welfare it might still not come about, since it could have negative consequences for one of the parties. Our paper differs from these contributions by considering the endogenous formation of an agreement that fulfills realistic contract restrictions and accounts for reasonable participation constraints of the contracting parties.

2 Salient features of investment agreements

The purpose of international investment agreements is to stimulate foreign investment by restricting host country incentives to take policy actions that reduce investor profits. The agreements occasionally provide rights regarding establishment, but the emphasis is on the treatment of foreign investment once in place.[9] The agreements should be distinguished from state-to-state tax treaties, although the latter might also importantly affect foreign investment flows. IIAs should in particular be distinguished from standard commercial contracts that are formed between a host country and an individual investor. There is currently no multilateral IIA, despite the attempt by the OECD to launch such an agreement in 1998; the World Trade Organization Agreement contains certain provision regarding trade-related investment, but no protection against direct or indirect expropriations. The very large number of investment agreements that are in force differ in coverage, and the associated case law is highly fragmented, with similar provisions interpreted very differently by different panels. But certain features are shared by the more prominent agreements, such as

[9] According to UNCTAD (2015, p.111), less than ten percent of IIAs include pre-establishment rights.

NAFTA, TPP, the recent mega-regional agreements, and the EU and US "model agreements."[10]

First, the agreements almost always request non-discriminatory treatment of foreign investment (and sometimes also investors), requesting host country treatment of foreign investment to be "no less favorable than that it accords, in like circumstances" to its own investors, so-called National Treatment, or to third country investors, so-called Most-Favored Nation Treatment.[11,12]

Second, investment agreements typically provide that foreign investment should be given at least a "minimum standard of treatment." A common part of these undertakings is a commitment to provide "fair and equitable treatment." There have been a number of contentious interpretations of this amorphous concept in case law.

Third, investment agreements almost invariably include rules concerning expropriation that not only cover direct expropriation, but also "...an action or series of actions by a Party [that] has an effect equivalent to direct expropriation without formal transfer of title or outright seizure"—so called "indirect expropriations." Expropriations are only allowed if they are for a public purpose, are non-discriminatory, are in accordance with due process of law, and if investors receive "prompt, adequate, and effective compensation." To provide some guidance for the interpretation of "indirect expropriation" agreements occasionally provide further specifications; for instance, whether an indirect expropriation is at hand depends on "the extent to which the government action interferes with distinct, reasonable investment-backed expectations" and "the character of the government action." Investment agreements also increasingly include restrictions on the ambit of the indirect expropriation clauses, so called *carve-outs.* For instance:

> "Nothing in this Chapter shall be construed to prevent a Party from adopting, maintaining or enforcing any measure otherwise consistent with this Chapter that it considers appropriate to ensure that investment activity in its territory is undertaken in a manner sensitive to environmental, health or other regulatory objectives."

and

> "Non-discriminatory regulatory actions by a Party that are designed and applied to protect legitimate public welfare objectives, such as public health, safety and the environment, do not constitute indirect expropriations, except in rare circumstances."

A standard specification concerning the required magnitude for compensation in case of expropriation is that it should be equivalent to the "fair market value" of the expropriated investment.

[10] See Dolzer and Schreuer (2012) for a comprehensive overview of international investment law.

[11] Unless otherwise stated, the quotations in this Section are taken from the US Model Bilateral Investment Treaty 2012. They appear verbatim in Chapter 9 of TPP, and to large extent also in Chapter 11 of NAFTA.

[12] There are significant exemptions from these requirements in some agreements for certain types of discriminatory measures, or for certain industries.

When interpreting this and related concepts, arbitration panels normally seek guidance in the general principles concerning state responsibility in customary international law. These hold that in case of illegal acts,

> "...reparation must, as far as possible, wipe out all the consequences of the illegal act and re-establish the situation which would, in all probability, have existed if that act had not been committed."[13]

and

> "[t]he compensation shall cover any financially assessable damage including loss of profits insofar as it is established."[14]

Arbitration panels have used a variety of ways to determine the fair market value, some forward-looking (such a discounted cash flows), and some backward-looking (incurred investment costs, for instance). Importantly for what follows, the purpose of the payment is to compensate the injured party for its losses, not to punish the responsible state: "A tribunal shall not award punitive damages."[15]

Finally, many investment agreements include a compulsory dispute settlement mechanism. These mechanisms fundamentally differ from those in trade agreements in several respects. For instance, they do not only allow state-to-state dispute settlement, they also enable foreign investors to litigate against host country governments (ISDS). Another difference is that the enforcement of rulings is much stronger than in trade agreements, since an investor often can request courts at home, in the host country, as well as in third countries, to seize assets belonging to the host country. We assume that the ISDS is sufficiently strong to render agreements perfectly enforceable.

As stated above, the purpose of the paper is to examine the design and implications of Pareto optimal investment agreements when constrained to be of the same form as actual agreements. We therefore assume that:

Assumption 1 *Feasible agreements have the following features:*
(1) Compensation is paid only in case of regulation;
(2) Compensation to each firm equals its operating profit;
(3) There are no payments to or from outside parties;
(4) Agreements impose no taxation or performance requirement on investors; and
(5) Investment decisions and regulation decisions are left at the discretion of investors and the host country, respectively.

[13] This often quoted passage is from the determination by the Permanent Court of International Justice in *The Factory at Chorzów* case 1928.

[14] Article 36, International Law Commission (2001). A footnote is omitted.

[15] See Crawford (2002, p. 219).

These restrictions ensure that feasible agreements cannot subsidize investment, nor impose punitive damages on host countries, since compensation cannot exceed foregone operating profits and can only be paid in case of regulation. This does not preclude other legal arrangements between host countries and individual firms or the industry, but these would then be subsumed in the domestic welfare and profit functions. The only deviation from these restrictions we will make is in Section 5.6, where we allow compensation to diverge from operating profits.

The five features listed above reflect the fact that investment agreements typically are long-term incomplete insurance contracts that cover a wide range of industries. This broad scope explains why compensation is paid only in case of regulation (Restriction 1), agreements impose no taxation or performance requirements on investors (Restriction 4), and why investment decisions and regulation decisions are decentralized (Restriction 5). When governments seek to provide more fine-tuned incentives for investment, this is done through commercial agreements with specific firms or industries. An important reason for why agreements do not rely on payments to and from outside parties (Restriction 3) is probably a lack of third party institutions willing to accept this role.

The most conspicuous assumption is that any compensation should equal foregone operating profits (Restriction 2). As mentioned above, the design of investment agreements in this regard reflects a basic principle in Customary International Law according to which state responsibility does not go beyond the restoration of the situation before the act. This principle is reflected for instance in the remedies available within the WTO. It would of course be possible for two countries to negotiate more general and non-linear compensation schemes that allow for payments to exceed foregone operating profits. But the purpose of such a scheme would be to fine-tune the investment incentives. This is the role of commercial contracts. Also, as we will show, simple compensation schemes are not always restrictive in terms of welfare.

3 The setting

Consider an industry in country i in which firms from country j can produce for the local market. The interaction in the industry occurs in stages. The foreign firms first make simultaneous irreversible investment in order to enter the industry. The host country is subsequently hit by an industry-specific exogenous shock that affects the country's benefit from production. Having observed the shock, the host country decides whether to allow production or to regulate by shutting the industry down completely. In the final stage there is production and consumption, if the industry is not regulated. We start by laying out the entirely standard final stage.

3.1 Product market competition

The representative consumer in country i maximizes a quasi-linear utility $\Omega^i(z_i) + z_0$ over the consumption z_i of the domestic good subject to the budget constraint $p_i z_i + z_0 \leq \Upsilon^i$, where p_i is the unit price of the domestic good, z_0 is a numeraire good, and Υ^i the exogenously given income. Ω^i is continuous, strictly increasing and strictly concave in the relevant domain, and $\Omega^i(0) = 0$. Income is sufficiently large that the consumer always purchases both goods in strictly positive amounts.

The total production cost of a representative foreign firm j is $C^j(x_i, k_i)$, where x_i is its production volume, and k_i its investment. The cost function has standard properties: $C^j(0, k_i) = 0$, $C^j_x > 0$, $C^j_{xx} > 0$, $C^j_k < 0, C^j_{kk} > 0$, $C^j_{xk} < 0$, and $C^j_{xx}C^j_{kk} \geq C^j_{xk}C^j_{kx}$.[16]

The market is for the most part assumed to be perfectly competitive, so the representative firm maximizes profit $p_i x_i - C^j(x_i, k_i)$ over production, taking the price in market i as given. The equilibrium output $X^i(k_i)$ and market-price $P^i(k_i)$ are in standard fashion defined by

$$\Omega^i_z(X^i(k_i)) = C^j_x(X^i(k_i), k_i) \text{ and } P^i(k_i) \equiv \Omega^i_z(X(k_i)). \tag{1}$$

with $X^i_k(k_i) > 0$, and $P^i_k(k_i) < 0$. To distinguish between the maximization problem facing the price-taking investors, and the problem of maximizing aggregate welfare, we define two reduced form expressions for operating profits:

$$\begin{aligned}\hat{\Pi}^j(p_i, k_i) &\equiv p_i X^i(k_i) - C^j(X^i(k_i), k_i) \\ \Pi^j(k_i) &\equiv \hat{\Pi}^j(P^i(k_i), k_i).\end{aligned}$$

In addition to generating local consumer surplus, the production causes a stochastic country-specific externality. The total welfare that the host country derives from production in the industry—its "domestic welfare"—is

$$S^i(k_i, \theta_i) \equiv \Omega^i(X^i(k_i)) - P^i(k_i)X^i(k_i) + \Psi^i(k_i, \theta_i).$$

where the first two terms represent conventional consumer surplus (disregarding the constant income Υ^i), and the last term the production externality. The magnitude of the externality depends on the stochastic shock $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$. The production externality Ψ^i can be either positive or negative, but a larger shock θ_i is defined to always correspond to a more negative externality: $\Psi^i_\theta < 0$. In contrast to much of the takings literature, we allow the externality to be a function of the investment, which for simplicity is concave: $\Psi^i_{kk} \leq 0$.

It is commonplace to refer to the benefits of foreign direct "investment" when pointing the positive effects of foreign firms in the local economy. But the alleged benefits mostly stem from the production, rather than the investment, by foreign firms. There is an important distinction between these concepts in the present analysis since we disregard from any effects of the investment as such on the local economy, and focus on effects emanating from production. Consequently, we refer to the term $\Psi^i(k_i, \theta_i)$ as capturing "production externalities."

The realizations of θ_i could represent the arrival of information concerning adverse environmental or heath consequences of the production process, as in the case of the Fukushima disaster, or concerning product characteristics, as in the case of tobacco. The model could also capture a

[16] Subscripts attached to function operators denote partial derivatives throughout.

situation where a host country has made implicit or explicit promises to pursue a certain policy, but where a financial shock induces the country to change its policy, such as in the case of the subsidies to renewable energy in Spain. We will not adopt any particular interpretation, but simply denote it as a "regulatory shock."

The distinction between the benefits from production in terms of local consumption, $\Omega^i(z_i) - p_i z_i$, and the production externalities $\Psi^i(k_i, \theta_i)$, serves to highlight the importance of whether the externalities are pecuniary or non-pecuniary. But it could justifiably be argued that the perceived benefits of foreign investment for the local economy are normally not in the form of increased product supply, but externalities in the form of employment, technological spill-overs, learning-by-doing by the work-force, etc. The model is fully compatible with the existence of such externalities. To illustrate, let $\Psi^i(k_i, \theta_i) \equiv \Lambda^i(X^i(k_i)) - \Phi^i(X^i, \theta_i)$, where $\Lambda^i(X^i(k_i))$ captures the externalities for the local economy from foreign production, and $\Phi^i(X^i, \theta_i)$ the adverse effects of the regulatory problem that stochastically affects the economy ($\Phi^i_\theta > 0$).

As mentioned, we allow for both positive and negative externalities ($\Psi^i(k_i, \theta_i) \gtreqless 0$). But we assume that the negative effects dominate for sufficiently large shocks, so that there are realization of θ_i for which the host country regulates absent an investment agreement. We also allow the marginal production externality to be negative ($\Psi^i_k < 0$). This could happen if production emits local pollution, whereas a substantial share of the local employment and business effects materialize during the construction phase instead of being related to production. If so, many of the positive host country externalities of foreign investment would be sunk at the production stage, although we for simplicity assume them to be zero.

3.2 Regulation

Host country i decides whether to allow production or regulate after observing θ_i. When making the regulatory decision, the host country *disregards* the consequences of its decisions for foreign investors. Allowing production yields domestic welfare $S^i(k_i, \theta_i)$, and foreign firms make the operating profits $\Pi^j(k_i)$. Since all externalities from investment are related to production, regulation by country i implies that $\Psi^i = 0$, that there is no consumption, and that operating profits are zero. Let the domestic welfare be strictly positive for any $k_i > 0$ if the regulatory shock equals $\underline{\theta}_i$, but very negative if the shock equals $\bar{\theta}_i$. Let $\Theta^i(k_i)$ be the regulatory shock for which the host country is indifferent between allowing production and regulating; it is given by

$$S^i(k_i, \Theta^i) \equiv 0. \tag{2}$$

Absent investment protection, the host country will regulate production if and only if $\theta_i > \Theta^i(k_i)$ since $S^i_\theta < 0$.

We will use the unweighted sum of welfare of the two countries that is generated in country i, denoted $W^i(k_i, \theta_i) \equiv S^i(k_i, \theta_i) + \Pi^j(k_i)$, as a benchmark to measure the extent to which the various regimes under consideration efficiently exploit the potential gains from cooperation. This is what a negotiated settlement would achieve if the countries were perfectly symmetric, or if they had access to side payments, and we therefore denote this "global welfare." The ex post globally optimal threshold for regulation $\Theta^{iG}(k_i)$ is thus given by

$$W^i(k_i,\Theta^{iG}) \equiv 0.$$

It follows from $\Pi^j(k_i) > 0$ and $S_\theta < 0$ that $\Theta^{iG}(k_i) > \Theta^i(k_i)$. Consequently:

Observation 1 *Absent investment protection, host country i regulates more frequently for any investment level k_i than what would maximize global welfare. There will be:*
(a) efficient production for $\theta_i \leq \Theta^i(k_i)$;
(b) overregulation from a global welfare point of view for $\theta_i \in (\Theta^i(k_i), \Theta^{iG}(k_i))$; and
(c) efficient regulation for $\theta_i \geq \Theta^{iG}(k_i)$.

3.3 Investment

Individual firms are sufficiently small to disregard their individual impacts on the probability of regulation and on the market price, but they rationally foresee the equilibrium levels of both. The expected profit of the representative firm is $F^i(\hat{\theta}_i)\hat{\Pi}^j(p_i, k_i) - R^j(k_i)$, where $\hat{\theta}_i$ is the threshold value for regulation, and where $R^j(k_i)$ is the investment costs; $R^j(0) = 0$, $R^j_k > 0$, and $R^j_{kk} \geq 0$. The associated first-order condition

$$-F^i(\hat{\theta}_i)C^j_k(X^i(K^i),K^i) - R^j_k(K^i) \equiv 0 \tag{3}$$

yields investment $\hat{k}_i$ as an increasing function $K^i(\hat{\theta}_i)$ of the expected cut-off level for regulation $\hat{\theta}_i$.

In contrast, the globally efficient level of investment maximizes

$$\int_{\underline{\theta}_i}^{\hat{\theta}_i} [S^i(k_i, \theta_i) + \Pi^j(k_i)]dF^i(\theta_i) - R^j(k_i). \tag{4}$$

The associated first-order condition

$$-F^i(\hat{\theta}_i)C^j_k(X^i(K^{iG}),K^{iG}) - R^j_k(K^{iG}) + \int_{\underline{\theta}_i}^{\hat{\theta}_i} \Psi^i_k(K^{iG},\theta_i)dF^i(\theta_i) \equiv 0 \tag{5}$$

gives the globally optimal investment level as a function $K^{iG}(\hat{\theta}_i)$ of the foreseen cut-off level for regulation $\hat{\theta}_i$. By a comparison of (3) and (5), and using $K^i_\theta > 0$:

$$K^i(\hat{\theta}_i) < K^{iG}(\hat{\theta}_i) \text{ iff } \int_{\underline{\theta}_i}^{\hat{\theta}_i} \Psi^i_k(K^{iG}(\hat{\theta}_i),\theta_i)dF^i(\theta_i) > 0.$$

Hence:

Observation 2 *Absent investment protection, firms underinvest for any regulatory threshold $\hat{\theta}_i$ relative to what would maximize global welfare, if and only if the marginal expected production externality is positive evaluated at the globally efficient level of investment.*

3.4 Equilibrium

Absent investment protection, investments will be chosen to maximize expected profit given the equilibrium threshold for regulation: $k_i^N = K^i(\theta_i^N)$. The threshold for regulation will in turn maximize the host country's ex post welfare given the investment: $\theta_i^N = \Theta^i(k_i^N)$. It follows that θ_i^N represents a Nash equilibrium if and only if $\theta_i^N = \Theta^i(K^i(\theta_i^N))$. To ensure the existence of a unique equilibrium, we assume throughout the analysis that if θ' solves $\theta = \Theta^i(K^i(\theta))$, then

$$\hat{\theta}_i < \Theta^i(K^i(\hat{\theta}_i)) \text{ iff } \hat{\theta}_i < \theta', \tag{6}$$

and we make the corresponding assumption regarding the function Θ^{iG}. These assumptions correspond to the "stability" conditions that are used in e.g. oligopoly models to rule out counter-intuitive comparative statics properties.[17]

Condition (6) has the intuitively appealing implication that the direct reduction in investment that results from the host country's disregard of foreign investor interests in its regulatory decision, does not indirectly induce the host country to reduce its regulation to the extent that there is in equilibrium *less* regulation than there would be if the host country took full account of foreign investors' profits. The latter outcome (k_i^E, θ_i^E) would be given by $k_i^E = K^i(\theta_i^E)$ and $\theta_i^E = \Theta^{iG}(k_i^E)$:

Observation 3 *Absent investment protection, the host country regulates more frequently, and firms invest less, than when regulation is ex post efficient.*[18]

The Observation highlights the implication of the host country disregard of the interests of foreign investors, assuming that the investments are market determined. But it does not necessarily follow that the there is too much regulation or too little investment relative to the globally efficient solution $k_i^G = K^{iG}(\theta_i^G)$ and $\theta_i^G = \Theta^{iG}(k_i^G)$. On the one hand, the tendency toward overregulation absent investment protection, captured formally by $\Theta^{iN}(k_i) < \Theta^{iG}(k_i)$, tends to dissuade investment. On the other hand, firms do not take into consideration the externality associated with production, which could give too much or too little investment depending on the sign of the production externality: $K^i(\hat{\theta}_i) \gtreqless K^{iG}(\hat{\theta}_i)$. We will analyze these distortions in more detail below.

4 Investment agreements

The previous Section laid out our modeling framework, and showed how the irreversibility of investment, combined with host country disregard of the interests of foreign investors, can cause investment to be too low from the point of view of both host and home countries. There is hence potentially scope for an investment agreement that insures foreign investors against losses from

[17] The stability condition (6) implies that there exists at most one solution θ'. Existence follows by way of the mean-value theorem, $\Theta^i(K^i(\underline{\theta}_i)) - \underline{\theta}_i \geq 0$ and $\Theta^i(K^i(\bar{\theta}_i)) - \bar{\theta}_i \leq 0$.

[18] If $\theta_i^E \leq \theta_i^N$, then $\theta_i^E \leq \Theta^i(k_i^E) < \Theta^{iG}(k_i^E) = \theta_i^E$, where assumption (6) implies the weak inequality and $\Theta^{iG}(k_i) > \Theta^i(k_i)$ the strict inequality. This is a contradiction, so $\theta_i^N < \theta_i^E$. Then, $k_i^N < k_i^E$ follows from $K_\theta^i > 0$.

regulation. A very simple agreement would request investors to receive full compensation for foregone operating profits whenever there is regulation. This would clearly solve the underinvestment problem. But it might instead induce foreign firms to overinvest, since foreign investors will attach weight to high realizations of θ_i for which their investment has no social value. Investors' incentives are also distorted by their neglect of the externalities from their investment for host countries, and this might provide another reason why there will be overinvestment. Simple as it is, this observation points to a fundamental feature of investment agreements that apparently is not fully appreciated by proponents of these agreements: regulatory regimes normally exist to addresses a potential *over*-investment problems. It is therefore inherently possible to get *too much* foreign investment if regulation is dissuaded through investment agreements. Indeed, an agreement that compensates for all regulation might reduce global welfare even relative to a no-agreement situation.[19] Consequently, an optimal investment agreement must have "carve-outs" from the compensation requirements, as also noted in the takings literature. But how should they optimally be design and what are the implications of such agreements?

4.1 Basic features of optimal compensation schemes

The following Proposition, which we prove in Appendix A.2, identifies properties of optimal investment agreements that fulfill restrictions (1)-(5) in Section 2 on the set of feasible agreements:

Proposition 1 *Any Pareto optimal investment agreement can be characterized in terms of a threshold $\hat{\theta}_i$ for each host country i, and a corresponding compensation rule*

$$T^i(k_i, \theta_i) = \begin{cases} \Pi^j(k_i) & \text{if } \theta_i \leq \hat{\theta}_i \\ 0 & \text{if } \theta_i > \hat{\theta}_i. \end{cases} \tag{7}$$

It is ex post optimal for host country i:
(a) to allow production if $\hat{\theta}_i \leq \Theta^i(k_i)$ and $\theta_i \leq \Theta^i(k_i)$;
(b) to allow production if $\hat{\theta}_i > \Theta^i(k_i)$ and $\theta_i \leq \min\{\hat{\theta}_i; \Theta^{iG}(k_i)\}$; and
(c) to regulate otherwise.

The first part of the Proposition thus demonstrates that for a given level of investment k_i, a compensation scheme that requests compensation if and only of the regulatory shock is below a certain threshold value $\hat{\theta}_i$, can be at least as good as *any* other scheme respecting restrictions (1)-(5). This holds regardless of how complex this latter scheme is with regard to the pattern of shocks for which it requests compensation. Henceforth, we refer to $\hat{\theta}_i$ as the *level of investment protection* in country i.

The second part of the Proposition characterizes the ex post regulation that will be implemented, for a given level of investment protection. It shows that the agreement will prevent regulation

[19] An example verifying this claim is available upon request. This particular moral hazard problem was discovered by Blume et al (1984) who were the first to discuss compensation schemes in the context of regulatory takings.

that would otherwise occur for $\theta_i \in (\Theta^i(k_i), \hat{\theta}_i]$, if $\hat{\theta}_i < \Theta^{iG}(k_i)$, and for $\theta_i \in (\Theta^i(k_i), \Theta^{iG}(k_i))$ if $\hat{\theta}_i \geq \Theta^{iG}(k_i)$. The agreement will hence have a bite for these realizations of θ_i. For the remaining realizations of θ_i there would either be no regulation absent an agreement, or there will be regulation in any event.

Note that Proposition 1 is valid for a range of frameworks other than the one laid out above. For instance, we show in the Appendix that the carve-out policy remains optimal also if we allow the agreement to employ the more general compensation scheme $T^i(k_i, \theta_i) = b_i(\theta_i)\Pi^i(k_i)$, where $b_i(\theta_i) \in [0, 1]$. It is also demonstrated that Proposition 1 holds for imperfect competition, mixed foreign/domestic ownership structures, firm asymmetries, and when firms take account of how their investment decisions affect the probability of regulation. In all these settings, the optimal agreement features a compensation scheme with the same structure as in (7).[20]

Proposition 1 has direct implications for one of the core claims in the policy debate, which holds that investment agreements cause *regulatory chill*. This notion is rarely precisely defined and can be given at least two different interpretations within the context of our model. An agreement can be said to cause *domestic* regulatory chill if the associated compensation scheme prevents host countries from undertaking regulations that they would have chosen in the absence of any agreement. Proposition 1 implies that domestic chill indeed will occur for shocks in the interval $\Theta^i(k_i) < \theta_i < \min\{\hat{\theta}_i; \Theta^{iG}(k_i)\}$. But this follows almost trivially, since the intended feature of the agreement is to compensate investors for losses from certain regulatory interventions, and the compensation payments will occasionally dissuade regulation. A corresponding *global* chill occurs if the agreement induces host countries to avoid regulation in situations in which such intervention would have been globally desirable. The Proposition shows that an optimal protection scheme *never features regulatory chill from a global perspective*: the fact that the compensation cannot exceed operating profits induces host countries to regulate the industry for all $\theta_i > \Theta^{iG}(k_i)$, regardless of whether regulation is compensable or not. Indeed, rather than causing global regulatory chill, agreements will result in ex post excessive or optimal regulation from a global point of view.

Observation 4 *An investment agreement will not cause global regulatory chill.*

[20] It could be objected that investment agreements typically do not specify country-specific compensation rules. This objection is correct as far as the texts are concerned. But most agreements allow for exceptions from the compensation requirements for certain types of regulations, for instance where the regulations cannot be considered as "manifestly excessive," or are "designed and applied" to protect legitimate public welfare objectives. Such a balancing of different regulatory motives clearly has to be done on a case-by-case basis, taking into account the specifics of the situation at hand, as also recognized in e.g. the TPP: "The determination of whether an action or series of actions by a Party, in a specific fact situation, constitutes an indirect expropriation, requires a case-by-case, fact-based inquiry..." (Annex 9-B.3(a) TPP). Compensation requirements are in this sense implicitly industry specific. Indeed, it is not even conceptually clear what it would mean to apply the same threshold across different industries. For instance, what does it mean to require that compensation should be paid for the same set of regulatory shocks in the oil refinery industry as for regulatory shocks in some pharmaceuptical industry?

It should be noted that the finding concerning global regulatory chill generalizes to a variety of settings, and holds for *all* schemes that feature non-negative compensation, not only those that compensate firms for lost operating profit. It can be found in its most general formulation as Theorem A.1 in Appendix A.2.

To see why global regulatory chill cannot occur under any optimal non-negative compensation mechanism, note that in order for to occur, the agreement must induce the host country to allow production for some realizations of θ_i for which it would have been better for the economy as a whole to regulate, that is, for some $\theta_i > \Theta^{iG}(k_i)$. This requires in turn that the host country is required to compensate investors with more than their foregone profits these realizations of θ_i. However, this cannot be optimal: if the compensation requirement would be reduced to $\Pi^j(k_i)$ for these shocks, it would become ex post optimal to regulate, but the investment incentive would remain unaffected because the profit would still be $\Pi^j(k_i)$. This modification would increase domestic regulatory efficiency without affecting investments and would therefore represent a domestic as well as a Pareto improvement.

The findings above have an immediate implication for the legal discussion concerning whether compensation should reflect incurred investment costs, foregone operating profits, or have some other base. From the point of view of ex post efficiency, it is desirable to get the host country to effectively internalize all ramifications of its regulatory decisions. Letting compensation be based on foregone operating profits serves to align host country interests with total welfare. Hence, for ex post efficiency, compensation should be based on, and be equal to, foregone operating profits.

Observation 5 *Host country and global ex post interests are aligned when compensation equals investors' foregone operating profits.*

4.2 The level of investment protection in investment agreements

Proposition 1 implies that we can characterize any international investment agreement between two countries in terms of two thresholds $(\hat{\theta}_1, \hat{\theta}_2)$ with associated equilibrium investment levels $(\hat{k}_1, \hat{k}_2)$, where $\hat{k}_i = K^i(\hat{\theta}_i)$. The building blocks for the analysis of the incentives to form investment agreements—expected domestic welfare and expected profits—are functions of those thresholds. This simplification is very useful considering the complexity of the agreements that would otherwise need to be considered.

The expected domestic welfare for host country i from an agreement requesting compensation in case of regulation with $\theta_i \leq \hat{\theta}_i$ equals

$$\tilde{S}^i(\hat{\theta}_i) \equiv \int_{\underline{\theta}_i}^{\min\{\hat{\theta}_i;\Theta^{iG}(\hat{k}_i)\}} S^i(\hat{k}_i, \theta_i)dF^i(\theta_i) - \Pi^j(\hat{k}_i)\int_{\min\{\hat{\theta}_i;\Theta^{iG}(\hat{k}_i)\}}^{\hat{\theta}_i} dF^i(\theta_i), \quad (8)$$

where the second term captures the expected costs to the host country for compensation payments in case the level of investment protection in country i (a level which is yet to be determined) exceeds the globally optimal level. The corresponding expected source country welfare of j equals the expected industry profit:

$$\begin{aligned}\tilde{\Pi}^j(\hat{\theta}_i) &\equiv \int_{\underline{\theta}_i}^{\min\{\hat{\theta}_i;\Theta^{iG}(\hat{k}_i)\}}\Pi^j(\hat{k}_i)dF^i(\theta_i) + \int_{\min\{\hat{\theta}_i;\Theta^{iG}(\hat{k}_i)\}}^{\hat{\theta}_i}\Pi^j(\hat{k}_i)dF^i(\theta_i) - R^j(\hat{k}_i) \\ &= F^i(\hat{\theta}_i)\Pi^j(\hat{k}_i) - R^j(\hat{k}_i).\end{aligned} \tag{9}$$

We assume throughout that the industry profit is increasing in the level of investment protection:

$$\tilde{\Pi}^j_\theta(\hat{\theta}_i) = \Pi^j(\hat{k}_i)f^i(\hat{\theta}_i) + X^i(\hat{k}_i)P^i_k(\hat{k}_i)K^i_\theta(\hat{\theta}_i)F^i(\hat{\theta}_i) > 0. \tag{10}$$

Intuitively, this assumption states that the direct effect of improved investment protection must dominate the indirect price effect of increased investments at the aggregate industry level. Assumption 10 is fundamental to our analysis and underlies many of our welfare and political economy results, as we shall see.

Let $\theta_i^U \in \arg\max_{\hat{\theta}_i} \tilde{S}^i(\hat{\theta}_i)$ be the maximal level of investment protection among those that maximize the expected domestic welfare in country i. By construction, the threshold θ_i^U is the upper bound to the level of domestic investment protection country i is willing to offer in an agreement where it has all the bargaining power and can make take-it-or-leave-it offers.

In principle, the expected investment externality could be so negative that country i would like to implement regulation for a range of $\theta_i < \theta_i^N$. But the only way a country could possibly enforce such a threshold would be if the industry was to *pay* compensation in case of regulation. As this possibility is ruled out, it is not ex post credible for the host country to regulate for any shocks $\hat{\theta}_i < \theta_i < \theta_i^N$. As we demonstrate in Appendix A.3:

Lemma 1 *The Nash equilibrium (θ_i^N, k_i^N) is the unique outcome of any agreement with a threshold $\hat{\theta}_i \leq \theta_i^N$.*

Lemma 1 implies $\theta_i^U \geq \theta_i^N$, but also $\hat{\theta}_i > \theta_i^N$ if country i is a host country. Obviously, the host country agrees on more investment protection, $\hat{\theta}_i > \theta_i^U$, than it would ideally prefer if $\theta_i^U = \theta_i^N$. But the two countries will never negotiate any level $\theta_i^N < \hat{\theta}_i < \theta_i^U$, either. The effects of changing $\hat{\theta}_i$ do not depend on $\hat{\theta}_j$ by the assumed separability of the two sectors. Source country j therefore would strictly benefit from increasing the level to θ_i^U from $\hat{\theta}_i$ by assumption (10). Host country i would not object to this change by the definition of θ_i^U. In fact, the source country, negotiating on behalf of its firms, would be able to use its bargaining power to increase investment protection even beyond this level.[21] We conclude that:

Proposition 2 *The negotiated level of investment protection will in any agreement exceed the level that maximizes domestic welfare in any host country i: $\hat{\theta}_i \geq \theta_i^U$, with strict equality if $\theta_i^U < \bar{\theta}_i$.*

[21] The marginal effect of an increase in $\hat{\theta}_i$ above θ_i^U is of first-order effect on industry profit, $\tilde{\Pi}^j_\theta(\theta_i^U) > 0$, but of second-order effect on domestic expected welfare, $\tilde{S}^i_\theta(\theta_i^U) = 0$, if $\theta_i^U \in (\theta_i^N, \bar{\theta}_i)$.

This finding that any host country typically allows an investment protection level above its ideal threshold applies to a much broader set of circumstances than the present setting because it only depends on our assumption that the expected industry profit is increasing in the level of investment protection.

4.3 North-South agreements

We will now draw on the above findings to examine the scope for investment agreements, and their distributional consequences. We use two stylized scenarios that highlight distinct rationales for investment agreements. The settings differ in two dimensions. First, countries have different abilities to make credible *unilateral* commitments to protect foreign investment. Developed countries are by and large able to implement such schemes through constitutions, laws and regulations, should they so desire. But developing countries typically do not have this capability for a variety of reasons; the credibility of incumbent governments might have been undermined by opportunistic regulations by previous governments, their political systems might be less stable, etc.. Second, the scenarios differ with regard to the *direction of the investment flows.* Again, there is a clear distinction between developed and developing countries, in that developed countries typically both are host and source countries for foreign investment, whereas developing countries primarily are host countries. We will henceforth distinguish between agreements that apply to unidirectional and bidirectional investment flows.

We first consider the traditional, and still most common type of investment agreement: a bilateral agreement between a developing and a developed country. While often formally symmetric, these are in practice highly asymmetric agreements since they are meant to remedy the developing country's inability to credibly protect foreign investment, thus effectively seeking to encourage increased investment flows from the developed to the developing country only. We denote this a "North-South" agreement.

The expected benefit from a North-South agreement for North (country j) is $\tilde{\Pi}^j(\hat{\theta}_i) - \tilde{\Pi}^j(\theta_i^N)$, since the alternative to the agreement is to have no investment protection at all. We have assumed that the expected industry profits increase with higher levels of investment protection, so North would benefit from any agreement with $\hat{\theta}_i > \theta_i^N$. A necessary and sufficient condition for there to be scope for a North-South agreement is therefore that South strictly benefits from a protection level above θ_i^N: $\tilde{S}^i(\hat{\theta}_i) > \tilde{S}^i(\theta_i^N)$ for some $\hat{\theta}_i > \theta_i^N$. This will be the case if the marginal consumer benefits, measured by the decrease in the consumer price of the good plus the expected marginal production externality, are positive:[22]

$$\tilde{S}_\theta^i(\theta_i^N) = -P_k^i(k_i^N)X^i(k_i^N)F^i(\theta_i^N) + \int_{\underline{\theta}_i}^{\theta_i^N} \Psi_k^i(k_i^N, \theta_i)K_\theta^i(\theta_i^N)dF^i(\theta_i) > 0. \tag{11}$$

But South might prefer no agreement to any enforceable level of investment protection: $\theta_i^U = \theta_i^N$.

[22] Since the market is perfectly competitive, there are no first order effects from either induced changes in production or investment levels.

An agreement might fail to arise even if there are investment protection levels $\hat{\theta}_i > \theta_i^N$ that would yield higher global welfare than no agreement. To see when this might arise, let

$$\tilde{W}^i(\hat{\theta}_i) \equiv \tilde{S}^i(\hat{\theta}_i) + \tilde{\Pi}^j(\hat{\theta}_i)$$

be the total expected (global) welfare generated in country i. If $\theta_i^U = \theta_i^N$, but

$$\tilde{W}_\theta^i(\theta_i^N) = \int_{\underline{\theta}_i}^{\theta_i^N} \Psi_k^i(k_i^N, \theta_i) dF^i(\theta_i) K_\theta^i(\theta_i^N) + \Pi^j(k_i^N) f^i(\theta_i^N) > 0, \tag{12}$$

then there would not be any agreement, even though this would increase global welfare. There are two first-order divergences between host country and global interests. First, the reduction in the consumer price that follows from increased investment protection —the first term in (11)—is only a transfer from firms to consumers from a global perspective. Second, the host country does not take into consideration the expected increase in operating profits of foreign investors from less frequent regulation—the second term in (12). An investment agreement might thus fail to form despite being desirable from a global perspective.[23]

If $\tilde{S}_\theta^i(\theta_i^N) > 0$, so that a North-South agreement is formed, it will feature an investment protection level θ^{NS} that reflects the relative bargaining strength of the two countries. North would choose the highest level of investment protection acceptable to South if in position to dictate the terms of the agreement, and South would choose θ_i^U if it had all the bargaining power. Since we assume that neither party can dictate the terms, the negotiated outcome θ^{NS} is strictly between these two levels (assuming $\theta_i^U < \bar{\theta}_i$). We conclude based upon these arguments and Proposition 3 that:

Proposition 3 *There need not be scope for a North-South agreement, even if such an agreement would increase expected global welfare. A North-South agreement will be formed if and only if it increases Southern domestic welfare, but it will entail more protection than is ex ante optimal for South.*

The welfare benefits from investment agreements in the North-South scenario stem entirely from the credibility it lends to South for compensating regulatory thresholds above θ_i^N. If South had full unilateral commitment possibilities, it would choose θ_i^U absent any agreement, in which case no agreement would be formed; $\tilde{S}^i(\hat{\theta}_i) - \tilde{S}^i(\theta_i^U) \leq 0$ for all $\hat{\theta}_i$. But since South lacks this commitment ability, it has to bargain with North over the level of investment protection, and will consequently have to accept a higher level of investment protection than what is optimal from the point of view of domestic welfare. Consequently, while an investment agreement can help South attract foreign investment from North, it is an imperfect substitute for credible domestic institutions from South's perspective.

Observation 6 *The rationale for a North-South agreement is South's lack of unilateral commitment possibilities regarding investment protection.*

[23] This case might arise even if there is underinvestment relative to first best investments, $k_i^N < k_i^G$, since the latter presume that it is possible to directly control investment.

4.4 North-North agreements

Consider next a scenario with an agreement between two developed countries. Such agreements typically differ from developed-developing country agreements in two respects. First, there are investment flows in both directions between the contracting parties. Second, countries are able to make credible unilateral commitments with regard to investment protection absent any agreement. To simplify matters we will focus on situations where the countries are mirror images in terms of demand structures, technologies, and propensity to experience regulatory shocks; this also seems broadly descriptive of the conditions facing the EU-US negotiations over TTIP. We will henceforth refer to this as a "North-North" agreement.

4.4.1 The rationale for North-North agreements

A North-North agreement will be symmetric by the symmetry of the two countries and the separability of the two sectors: $\theta_1^{NN} = \theta_2^{NN} = \theta^{NN}$. Hence, the expected welfare of country i equals $\tilde{S}^i(\theta^{NN}) + \tilde{\Pi}^i(\theta^{NN})$ under a North-North agreement. And since Northern countries are able to make unilateral commitments with regard to compensation, country i will not accept any agreement that yields less than $\tilde{S}^i(\theta^U) + \tilde{\Pi}^i(\theta^U)$, where the level θ^U of investment protection that maximizes the expected domestic welfare is the same for both countries. By industry symmetry, $\tilde{\Pi}^i(\hat{\theta}) = \tilde{\Pi}^j(\hat{\theta})$, and therefore any agreement satisfies $\tilde{W}^i(\theta^{NN}) > \tilde{W}^i(\theta^U)$. Hence, a North-North agreement will arise if and only if it increases global welfare. The problem in North-South relations, where globally welfare improving agreements might not occur, does not arise in a North-North setting. The symmetric bidirectional investment flows cause both countries to effectively internalize the profit of FDI flowing into the country. A sufficient condition for there be scope for an agreement therefore is $\tilde{W}_\theta^i(\theta^N) > 0$ even if $\theta^U = \theta^N$.[24]

Proposition 4 *There is scope for a North-North investment agreement if and only if such an agreement increases expected global welfare.*

The North-South and North-North scenarios identify two separate sources of gains from an investment agreement. The North-North scenario illustrates the bargaining gains to be had from the *internalization of negative international externalities* associated with national regulatory policies. Since investments flow in both directions, the two countries can negotiate improved investment protection abroad by offering improved investment protection at home. This corresponds closely to the standard view of the gains from trade agreements, according to which these agreements are means to taking countries out of Prisoners' Dilemmas, allowing them to exchange increased

[24] One could argue that Northern countries should not be constrained to compensation schemes fulfilling restrictions (1)-(5) when making unilateral commitments to investment protection. If so, this would reduce the scope for an investment agreement. But since host countries still do not internalize the effects of their protection for foreign investors, there might still be room for a welfare improving agreement of the type we are considering; see Proposition 5.

imports against increased exports to the mutual benefit of all. The North-South scenario instead highlights how investment agreements enable countries to enter into credible investment protection commitments that they otherwise would not be able to take on. Such a mechanism corresponds closely to the "commitment approach" to explaining trade agreements, which sees these agreements as helping governments to withstand domestic protectionist pressures.[25]

Observation 7 *A North-North agreement solves a Prisoners' Dilemma-like problem between the countries.*

Yet another difference between North-North and North-South agreements is the extent to which they could be improved upon through commercial contracts between host countries and individual investors. Such contracts would have the advantage of allowing the specification of investment levels, and possibly also regulatory policies. Commercial contracts are in this sense superior to state-to-state agreements for Southern countries, provided they can be formed without transaction costs (that is, disregarding cost of contracting costs, informational problems, etc.). Such commercial contracts cannot replace North-North agreements however, since the negotiations over these commercial contracts would not include the treatment of host country investors in the other country.

Observation 8 *Contracting on investment and regulation with individual investors would be better than a North-South agreement, but not necessarily a North-North agreement.*

4.4.2 The outcome of North-North agreements

Because of the country symmetries, it seems reasonable to assume that negotiation over a North-North agreement effectively maximizes global welfare: $\theta^{NN} = \arg\max_{\hat{\theta}} \tilde{W}(\hat{\theta})$. The global optimum also is the outcome of a North-South negotiation if the relative bargaining strength of the two countries is proportional to the relative net benefit of an agreement (and of course more generally to agreements that are negotiated with access to side payments).[26]

To identify any second-best welfare distortions involved with the determination of θ^{NN}, it is helpful to decompose the total investment distortion $k^G - k^N$ absent any agreement into two parts. To this end, recall that the threshold $\theta^E = \Theta^G(k^E)$ and the investment level $k^E = K(\theta^E)$ jointly represent the equilibrium absent an agreement if the host country were to fully internalize all effects on foreign profit in its regulatory decision, but firms invest to maximize expected profit. In this

[25] Bown and Horn (2015) informally discuss a similar distinction between traditional developing/developed country investment agreements, and investment agreements between developed countries. They suggest that the latter might not serve to address hold-up problems at all, but other forms of externality problems.

[26] Assume that $\hat{\theta}^{NS}$ is chosen to maximize the Nash product $[\tilde{S}^i(\hat{\theta}) - \tilde{S}^i(\theta^N)]^{\rho_i}[\tilde{\Pi}^j(\hat{\theta}) - \tilde{\Pi}^j(\theta^N)]^{\rho_j}$, where ρ_i (ρ_j) is the bargaining strength of South (North). Using $\tilde{S}^i_\theta(\theta^{NN}) + \tilde{\Pi}^j_\theta(\theta^{NN}) = 0$, it is straightforward to verify that $\theta^{NS} = \theta^{NN}$ if

$$\frac{\rho_i}{\rho_j} = \frac{\tilde{S}^i(\theta^{NN}) - \tilde{S}^i(\theta^N)}{\tilde{\Pi}^j(\theta^{NN}) - \tilde{\Pi}^j(\theta^N)}.$$

case, $k^E - k^N > 0$ is the distortion associated with the ex post incentive to regulate; see Observation 3. The remaining part, $k^G - k^E \gtrless 0$, is ambiguous and occurs because profit maximizing firms fail to account for the production externality. To simplify the exposition in this part of the analysis, we assume that the global welfare function is well-behaved:

$$\frac{d^2}{dk^2}[\int_{\underline{\theta}}^{\Theta^G(k)}(S(k,\theta)+\Pi(k))dF(\theta)-R(k)]<0,\ k>0. \tag{13}$$

North-North agreements yield three types of outcomes. The first arises when the marginal production externality is negative:

$$\frac{\tilde{W}_\theta(\theta^E)}{K_\theta(\theta^E)}=\int_{\underline{\theta}}^{\theta^E}\Psi_k(k^E,\theta)dF(\theta)<0. \tag{14}$$

This is equivalent to assuming that firms' profit maximizing behavior yields overinvestment in the sense that $k^E > k^G$. A reduction in the level of investment protection below θ^E reduces the problem of overinvestment, but implies ex post overregulation. The equilibrium protection level θ^{NN} and investment $k^{NN} \equiv K(\theta^{NN})$ investment solves $\tilde{W}_\theta(\theta^{NN}) = 0$ and strikes a balance

$$\Pi(k^{NN})f(\theta^{NN})=-\int_{\underline{\theta}}^{\theta^{NN}}\Psi_k(k^{NN},\theta)dF(\theta)K_\theta(\theta^{NN})-S^i(k^{NN},\theta^{NN})f(\theta^{NN}) \tag{15}$$

between the marginal benefit to the foreign firms of investment protection (the term on the left-hand side of the above equation) against the marginal production externality (the first term on the right-hand side) plus the cost of domestic underregulation (the second term on the right-hand side).[27]

The second kind of outcome arises when the marginal production externality is non-negative, $\tilde{W}_\theta(\theta^E) \geq 0$, but not too large,

$$\frac{\tilde{W}_\theta(\bar{\theta})}{K_\theta(\bar{\theta})}=\int_{\underline{\theta}}^{\Theta^G(\bar{k})}\Psi_k(\bar{k},\theta)dF(\theta)\leq(1-F(\Theta^G(\bar{k}))R_k(\bar{k}), \tag{16}$$

where $\bar{k} = K(\bar{\theta})$ is the investment level that results if foreign investors are fully compensation for any regulation. The profit maximizing behavior yields underinvestment under these assumptions: $k^E \leq k^G$.

The final equilibrium constellation arises when there is a large positive externality in the sense that (16) is reversed. There will then be full protection of operating profits, $\theta^{NN} = \bar{\theta}$, and under-investment.

We summarize the above as follows, which applies a fortiori to North-North agreements (the proof is in Appendix A.5):

[27] Alternatively, we can think of the first cost as the *intensive margin* of the production externality (firms overinvest from the host country viewpoint), whereas the second cost is the *extensive margin* of the production externality (there are too many firms from the host country viewpoint).

Proposition 5 *An agreement that maximizes the expected global welfare yields:*
(a) overinvestment and ex post overregulation, if the marginal production externality is negative;
(b) globally efficient investment and regulation if the marginal production externality is non-negative, but not too large;
(c) underinvestment and ex post optimal regulation if the marginal production externality is positive and large.

The Proposition establishes that a simple carve-out policy is capable of correcting both the regulatory *and* the investment distortion by means of a *single* policy variable in a robust set of circumstances, i.e., when the marginal production externality is positive. An increase in the level of investment protection up to the level θ^E improves firms' ex ante incentives to invest and reduces the host country's ex post incentive to regulate. Above this threshold, only the investment externality is left, and there exists a $\theta^{NN} > \theta^E$ that yields full efficiency (θ^G, k^G).[28]

Note that globally optimal agreement entails compensation payments in the (b) and (c) equilibria in the Proposition. The purpose of compensation payments is to improve investment incentives and is efficient in our setting. Equilibrium payments arise also in the setting employed by Aisbett et al (2010a), but in their case the agreement overcompensates firms for their losses because the regulatory shock is imperfectly observable and the host country would overregulate otherwise. Here, the payments arise even with full observability of the regulatory shocks.

Observation 9 *An agreement that maximizes global welfare will give rise to compensation payments in equilibrium when the marginal production externality is positive.*

4.4.3 Winners and losers from North-North agreements

As we have seen, a North-North agreement can partly or fully remedy the underinvestment and the overregulation problems, as long as there is scope for an agreement at all. But the gains will come with pronounced distributional implications: Northern countries can unilaterally implement their preferred protection level for their incoming investment. Each country therefore suffers a loss $\tilde{S}^i(\theta^U) - \tilde{S}^i(\theta^{NN}) > 0$ in domestic welfare by signing a North-North agreement. Country i will therefore sign such an agreement only if it is compensated for its domestic loss in terms of better investment protection abroad for its domestically-owned industry's FDI: $\theta^{NN} > \theta^U_j$.[29]

[28] In comparison, Blume et al (1984), Miceli and Segerson (1994) and Aisbett et al (2010a) establish efficiency in the knife-edge case when the marginal production externality is zero ($\Psi_k = 0$, $\forall\theta$) so that only the host country's incentive to regulate is distorted.

[29] Our finding that $\hat{\theta}^{NN} > \theta^U_i$ is implied by the assumption that the outside option is (θ^U_i, θ^U_j), but is not a necessary condition. The result arises also from Pareto efficient bargaining if we assume that both countries have bargaining power. We showed in the proof of Proposition 3 that $\hat{\theta}_i \geq \theta^U_i$ for any Pareto optimal agreement $(\hat{\theta}_i, \hat{\theta}_j)$, with strict inequality if $\hat{\theta}^U_i < \bar{\theta}_i$ and country j has bargaining power.

Proposition 6 *A North-North agreement benefits foreign investors in both countries, but reduces expected domestic welfare in both countries.*

Hence, the entire surplus from the North-North agreement can be attributed to an increase in industry profits because the unilateral protection level $\hat{\theta}_i^U$ already internalizes *all* domestic benefits of foreign investment for country i, even those associated with employment, local business development and so forth. Effectively, each party to a North-North agreement sacrifices domestic welfare to enhance the protection of its industry located abroad. Consequently, an alternative view of the purpose of an North-North agreement is the following:

Observation 10 *A North-North agreement is a means of exchanging reduced domestic welfare against improved protection of outgoing investment.*

The costs and benefits for the Southern party to North-South investment agreements have been discussed for years. But generally speaking, several thousands of such agreements have been signed, and very few have been terminated or not been renewed, without much political upheaval. This contrast sharply with the heated debate concerning the recent attempts to include investment protection in more symmetric agreements, and most notably in CETA, TPP, and TTIP. Our North-South and North-North scenarios point to a possible explanation for the much more critical public view of the mega-regional agreements: existing unilateral commitments by the EU and the US to protect incoming FDI already provide significant protection, and the *additional* investment protection that e.g. TTIP would offer mainly benefits foreign investors, and harms the rest of society. Incidentally, this distribution of the benefits of TTIP appears closely compatible with arguments put forward by the U.S. Administration and the EU Commission as to the benefits of investment protection in TTIP. Both sides rarely emphasize gains from increased flows of incoming investment, but mainly point to the benefits from increased protection of their respective *outward* flows.

5 Extensions

This section extends the above analysis in a number of directions. We consider political risks, non-discrimination clauses and direct expropriation. Then we examine the robustness of the optimal compensation rules to assumptions of market structure and asymmetric information. Finally, we derive a new compensation scheme based upon relative performance that can implement the global optimum under a range of circumstances. For simplicity, we assume that both countries are symmetric so that we can discard the country index i.

5.1 Political risks

A central issue in the policy discussion is the *nature of the shocks* that investment agreements should protect. In particular, a discussion evolves around the extent to which newly elected governments should be constrained by agreements they inherit from previous governments. A natural distinction can be made between regulation that occurs as a consequence of exogenous events versus interventions that result from an apparent changes in political preferences. An example of an exogenous regulatory shock is the Fukushima disaster in 2011 that lead to a global reassessment of the dangers of commercial nuclear power and caused Germany to accelerate its nuclear phase-out, and the wave of nationalization that occurred in Venezuela after Hugo Chavez came to power could exemplify political risks; both events have triggered litigation under investment agreements. Should regulatory and political shocks be compensated differently? On the one hand, critics of investment treaties sometimes argue that it is undemocratic to bind governments by undertakings entered into by previous governments to protect foreign investor interests. On the other hand, the consequences for the investors are the same no matter the reasons for why they are regulated.

To shed some light on economic considerations surrounding regulatory versus political uncertainty, we make a simple reformulation of the model above by assuming that the total shock is the product of two parts, $\theta = \eta\lambda$. The parameter η represents a regulatory shock, e.g. a reassessment of the dangers of nuclear power, whereas λ represents a political shock, e.g. a change in the preferences for nuclear power. What matters for our argument is that the total shock is captured by the parameter θ, multiplicity is only for simplicity. Assume that both shocks are resolved simultaneously between the investment stage and the production stage. If we let $F(\theta)$ be the cumulative distribution of the total shock in the host country, then the level of investment protection θ^{NN} characterized in Proposition 5 is part of a (globally welfare maximizing) North-North agreement. In particular, regulation for all $\eta\lambda > \theta^{NN}$ should leave firms without compensation independently of whether regulation occurs as a consequence of an exogenous regulatory shock (high η) or because of a change in political preferences (high λ); the two are perfect substitutes. Equivalently, the threshold θ^{NN}/λ for which a regulatory shock η represents a basis for legitimate public intervention is smaller when there is a stronger public preference for regulation (λ is higher). This representation of political preferences is admittedly of very reduced form, but nevertheless illuminates a key property of optimal compensation rules for regulatory interventions:

Observation 11 *The compensation to firms should depend on political preferences in an optimal investment agreement. With the total shock given by $\theta = \eta\lambda$, a government that is more sensitive to the regulatory shock should be allowed to regulate without compensation for a larger range of regulatory shocks.*

A well-functioning compensation scheme that incorporates the strength of political preferences as a basis for compensation requires that such preferences are verifiable and their magnitude quantifiable. Because intervention is ex post optimal for a broad range of shocks and political preferences, there is an obvious temptation for decision makers to attribute any intervention to a sufficiently strong change in political preferences to vindicate a compensation carve-out. A carve-out policy relying upon on λ would not be incentive compatible under asymmetric information about political preferences, an issue we discuss in more detail in Section 5.5.

Our finding that optimal compensation schemes depend on the future realization λ of political preferences builds upon an assumption that the compensation scheme is designed behind the veil of ignorance. In reality, the properties of the scheme is likely to reflect the political preferences of the government in place when the scheme is decided. The current government can bind the regulatory decisions of future governments independently of their political preferences by establishing that regulation should depend entirely upon η instead of $\eta\lambda$. Hence, there can be political economy reasons for why political preference shocks would be left out of investment agreements, although including them would be efficient.

5.2 National Treatment

Investment agreements typically include non-discrimination clauses; as briefly discussed in Section 2. We have so far steered away from discrimination against foreign firms, and thus from a role for a National Treatment (NT) clause, simply by assuming that there are no domestic firms. This is not quite as restrictive as it might seem, since there are in practice many instances where NT does not have a bite due to a lack of domestic firms that produce under sufficiently "like circumstances" to those facing foreign investors. To shed some light on the role of NT in our setting, we now assume that in each host country there is also a domestically-owned industry (indicated by subscript D) in addition to the foreign-owned industry (indicated by subscript F). The two industries are identical in terms of demand and production structures and suffer from the same country-specific shock, and for the purpose of an NT provision thus produce under "like circumstances." But the sectors are economically unrelated to one another to avoid strategic considerations affecting the regulation decisions. Each host country fully internalizes the consequences of regulation for the profits of its domestic industry, but continues to disregard the impact on foreign profits.

Assume first that there is no investment protection. Although the two industries are symmetric, they differ in terms of equilibrium investment because the host country will regulate the foreign industry more frequently than the domestic industry. The equilibrium is thus characterized by the thresholds $\theta_F^N = \theta^N$ and $\theta_D^N = \theta^E > \theta^N$, with the corresponding investment levels $k_F^N = k^N$ and $k_D^N = k^E > k^N$; see Observation 3.

Consider next the polar case of full unilateral commitment ability. Let $\theta^W \equiv \arg\max_{\hat{\theta}} \tilde{W}(\hat{\theta})$ be the protection level that maximizes the *expected* global welfare generated in the host country; recall that θ^E is what maximizes global welfare ex post. The host country will in this case choose carve-out policies with different levels of investment protection in the two sectors: $\theta_F^U = \theta^U$ and $\theta_D^U = \theta^W \geq \theta^U$ followed by equilibrium investments $k_F^U = K(\theta^U)$ and $k_D^U = K(\theta^W) \geq k^U$. These inequalities are strict in the typical case of incomplete investment protection ($\theta^U < \bar{\theta}$).[30]

Foreign firms can here be said to face two forms of discrimination. First, in both regimes

[30] By definition, θ^W equals the negotiated outcome θ^{NN} between two symmetric countries; see Proposition 5. From Proposition 2, we know that any negotiated outcome has a level of investment protection strictly above θ^U in any host country if $\theta^U < \bar{\theta}$.

the host country regulates foreign investment more frequently, $\theta_F^N < \theta_D^N$ and $\theta_F^U < \theta_D^U$, despite both industries being subject to the same shock θ. Second, the host country applies a stricter rule for when to regulate industry F for any investment level k, $\Theta(k) < \Theta^G(k)$. But such rule-based discrimination has no separate implications if individual firms treat decisions to regulate as unrelated to their own investment choices.

NT rules do not rule out more favorable treatment of the foreign investment. We will therefore represent an NT rule with the requirement that $\theta_F \geq \theta_D$. This restriction ensures that the domestic industry is regulated whenever there is regulation of the foreign industry, but the opposite need not hold.

5.2.1 Symmetric investment flows

Let us first examine the consequences of NT in the North-North scenario in Section **??**, where investments flow symmetrically between the two countries and countries can unilaterally commit to any level of investment protection. Symmetry across countries implies that they will agree on the same efficient protection level in their F sectors under an investment agreement. By symmetry across industries, this is exactly the same level of investment protection as in the domestic industries absent a NT clause: $\theta_D = \theta_F = \theta^{NN} = \theta^W$.

Contrast this scenario with an alternative setting with symmetric investment flows, but where countries are *cannot* make credible unilateral commitments; such as a South-South scenario. An efficient negotiation over investment protection for foreign investment absent NT would again lead to the protection level θ^W in the F sectors in the two countries. But assuming that countries cannot unilaterally commit to investment protection even for its domestic industry, they will apply a different threshold $\theta_D^N = \theta^E \neq \theta^W$ for regulation in the domestically-owned industry D. Adding NT will still be inconsequential if the marginal production externality as defined in Proposition 5 is non-negative, because then the foreign sector would enjoy (weakly) more protection than the domestic sector under the initial agreement: $\theta^W \geq \theta^E$. However, in the opposite case of a negative marginal expected production externality, the NT clause would enable countries to credibly reduce underregulation in their domestic sectors from θ^E to θ^W. This reduction in the level of investment protection represents a domestic welfare improvement that hurts the domestic industry less than it benefits the rest of the domestic economy:[31]

Proposition 7 *If investment flows are symmetric, then an NT clause*
(i) has no impact if countries can unilaterally commit to investment protection;
(ii) reduces investment protection of the domestic sector, thus hurting the domestically-owned industry, but increasing domestic welfare at large, if no country can unilaterally commit to investment protection.

[31] The protection level θ^W maximizes $\tilde{W}(\hat{\theta})$, and $\theta^E > \theta^W$ implies $\tilde{\Pi}(\theta^E) > \tilde{\Pi}(\theta^W)$. Hence, $\tilde{S}(\theta^W) - \tilde{S}(\theta^E) = \tilde{W}(\theta^W) - \tilde{W}(\theta^E) + \tilde{\Pi}(\theta^E) - \tilde{\Pi}(\theta^W) > 0$.

It is instructive to compare the role of NT here with its role in trade agreements. A main problem for trade agreements is that there is a myriad of domestic policies that can be used to undermine commitments concerning border instruments such as negotiated tariff reductions. For instance, a tariff binding could easily be rendered useless by the introduction of a "sales tax" on the imported product. The basic purpose of NT is to dissuade countries from thwart the agreements in such a fashion. NT seeks to achieve this by effectively forcing the importing country to distort its domestic sector if it wants to e.g. tax imports.[32] The purpose of NT in the present context is different however, since it is not meant to neutralize opportunistic behavior. Instead, NT here essentially serves to extend to the domestic sector the commitment possibilities that the investment agreement brings. This explanation for non-discrimination clauses seems more consistent with the commitment approach main to trade agreements, which sees such treaties as means of influencing the government's interaction with domestic constituencies.

Observation 12 *NT can allow countries that lack credible unilateral commitment possibilities to indirectly use the enforcement mechanism offered by investment agreements to solve underregulation problems in their domestic sectors.*

5.2.2 Unidirectional investment flows

The scenario with symmetric countries, sectors and investment flows is analytically very convenient as we can then reasonably focus on agreements that maximize global welfare. Matters are also simple in the case where investment flows in one direction only if the host country has the ability to unilaterally commit to investment protection. It can then do no better than to implement the domestically optimal thresholds $\theta_D^U = \theta^W$ and $\theta_F^U = \theta^U$. It therefore has no incentive to enter into any agreement with or without an NT provision.

Assume instead that the host country has no such commitment possibilities, as in the North-South scenario. An NT clause would again be redundant if the agreement would result in more investment protection for the foreign than the domestic industry $\theta^{NS} \geq \theta^E$, since the NT clause would then not have any bite. Hence, NT clauses are of economic importance with unidirectional investment flows only if host countries cannot unilaterally commit to investment protection, and if additionally an agreement without NT would imply more regulation of the foreign than the domestic industry: $\theta^{NS} < \theta^E$. To characterize the impact of an NT clause in such a scenario, we introduce the following (clumsy) notations. First, let θ_{NT}^U be the level of investment protection that maximizes expected host country welfare under NT. Second, let θ_{NT}^{NS} be the negotiated level of investment protection in a Pareto optimal North-South agreement under NT. We prove in Appendix A.4 that a national treatment clause increases the optimal level of investment protection from a domestic viewpoint:

Lemma 2 $\theta^U \leq \theta_{NT}^U \leq \theta^W$ *and* $\theta_{NT}^U \leq \theta_{NT}^{NS}$. *The inequalities are strict if* $\theta_{NT}^U \in (\theta^N, \bar{\theta})$.

[32] Horn (2006) examines the pros and cons of NT from this perspective.

Adding an NT clause to a North-South agreement will have economic implications only if South has a substantial degree of bargaining power, because North, who wants to maximize investment protection, would otherwise be able to negotiate a level of investment protection beyond the level that the host country is able to offer to the domestic industry (a level $\theta^{NS} \geq \theta^{E}$). If instead South has most of the bargaining power, then the negotiated outcome will be close to θ^{U} absent NT, and in particular $\theta^{U} < \theta^{NS} \leq \theta^{U}_{NT} < \theta^{NS}_{NT}$. In such an instance, North is strictly better better off by including an NT clause in the agreement since this will yield a higher level of investment protection ($\theta^{NS}_{NT} > \theta^{NS}$). The effect on the host country of the NT clause is ambiguous, however. The domestic welfare in sector D can potentially increase if the equilibrium level of investment protection is closer to the unilaterally optimal level θ^{W} than before. However, the domestic welfare is likely to fall in sector F because here an additional NT clause increases the level of investment protection even further beyond the domestically optimal level θ^{U}. We summarize these findings as:

Proposition 8 *If investment flows are unidirectional, then an NT clause*
(i) is irrelevant if countries can unilaterally commit to investment protection (since there will be no agreement), or if the source country has sufficient bargaining power to achieve $\theta^{NS} \geq \theta^{E}$*;*
(ii) benefits foreign investors, and has ambiguous effects on the host country, if the host country has sufficient bargaining power to achieve $\theta^{NS} \leq \theta^{U}_{NT}$.

5.2.3 Remarks

We have examined the implications of adding a non-discrimination clause to an agreement that negotiates the level of protection for foreign investment. It is a common view among critics of investment agreements that their *only* role should be to prevent unequal treatment of foreign investors. For instance, Stiglitz (2008, p. 249) argues in this vein that "...non-discrimination provisions will provide much of the security that investors need without compromising the ability of democratic governments to conduct their business." Restricting international investment clauses to merely include non-discrimination would increase (decrease) protection of the foreign (domestic) investments, with ambiguous consequences for the rest of society.[33] But abstaining from negotiating the investment protection levels for the foreign industry would normally reduce global welfare, since these negotiations internalize positive international externalities from protection schemes for foreign investment. In our setting, it might actually be better from a global welfare perspective to have no agreement at all compared to an agreement that only imposes NT.

[33] Formally, with a binding NT regime, and absent possibilities for unilateral commitments regarding protection levels, the common unilaterally determined protection level θ^{NT} will be given by

$$2S(K^{N}(\theta^{NT}), \theta^{NT}) + \Pi(K^{N}(\theta^{NT})) \equiv 0.$$

It is easy to verify that $\theta^{NT} \in (\theta^{N}_{F}, \theta^{N}_{D})$. With commitment possibilities, the unilaterally determined protection level will be given by θ^{U}_{NT} characterized in Lemma 2.

5.3 Direct expropriation

A primary objective of investment agreements is of course to prevent direct expropriation. Such instances now are less common than during the 1960s and 1970s, but not completely something of the past, as shown by Haizler (2012). IIAs typically have stricter rules regarding compensation for direct relative to indirect expropriation, in the sense that the exceptions described in Section 2, related for instance to environmental policy, only apply to the second type of expropriation. This stricter attitude might seem intuitively appealing since direct expropriations are (at best) pure transfers of rents. Matters are not quite as simple from a contractual point of view, however.

There are important distinctions between the two forms of expropriation. First, regulatory expropriations shut down production in the regulated entities, whereas they can continue in the case of a direct expropriation, albeit perhaps less efficiently operated by the government. Second, host countries benefit from regulatory expropriations by avoiding regulatory problems, but in the process suffer other welfare losses. With direct expropriation the benefits instead come from gaining access to a stream of operating profits. Surprisingly perhaps, these differences imply that it can actually be efficiency-enhancing to allow for direct expropriations.

Consider an initial investment agreement that yields equilibrium investment $\hat{k}$. Assume that there is inefficient overregulation for $\theta \in (\hat{\theta}, \Theta^G(\hat{k}))$. Consider a modified version with the same compensation rule for intervention as before, but now allows the country to directly expropriate for shocks in the domain $(\hat{\theta}, \Theta^G(\hat{k})]$. The additional benefit of the expropriated profit now will cause the host country to take over the firm and allow production for all shock realizations $\theta \in (\hat{\theta}, \Theta^G(\hat{k})]$ for which it previously shut down regulation. From the viewpoint of investors, it does not matter whether they are regulated or expropriated. They still receive the same treatment and compensation and therefore continue to invest $\hat{k}$. An agreement with regulation and expropriation therefore is more efficient than one without any possibilities for direct expropriation because production is ex post globally optimal for all $\theta < \Theta^G(\hat{k})$. Intuitively, direct expropriation can represent a more efficient means of preventing overinvestment than regulatory expropriation, because the former does not entail shutting down ex post valuable production. These beneficial features of direct expropriation can actually take us very far (the proof is found in Appendix A.6):

Proposition 9 *The global welfare optimum (k^G, θ^G) can be implemented as a Nash equilibrium through an international investment agreement that allows direct expropriation under the assumption that expropriation does not reduce the profits of the expropriated assets.*

The point here is not to argue that direct expropriations necessarily should be allowed in actual agreements, but rather that the reason for not doing so is not as trivial as it might seem. The common negative perception of direct expropriations is probably often based on the notion that they effectively constitute unproductive (or worse) thefts that deter investment. But it is exactly the fact that they constitute a pure transfer of ownership that might provide a role for them in investment agreements to mitigate the overinvestment problem stemming from the full compensation used to reduce the ex post incentives to regulate.

An important caveat for efficiency enhancing direct expropriation is that the shock θ has to be observable and verifiable for a globally efficient outcome to be achievable. Assume instead that the host country has private information about θ. The value of allowing production in the foreign-owned industry is $S(k,\theta)$, whereas the value of direct expropriation is $S(k,\theta)+\Pi(k)-T^x$, where T^x is what the host country must pay in compensation to foreign investors under direct expropriation. The net benefit $\Pi(k) - T^x$ of direct expropriation is independent of the realization of θ. The host country would never truthfully reveal θ if the expropriation compensation depended on the shock. Hence, the only way an IIA can ensure foreign ownership in the host country ever to be profitable is to set $T^x \geq \Pi(k)$. With this compensation scheme, the host country either allows private production or regulates, but direct expropriation can never be strictly beneficial to the host country. If the above inequality was reversed, then the host country would always intervene in the market, either by direct expropriation or through regulation. But it would never be optimal for the host country to maintain private ownership of the foreign industry.

We conclude that it impossible to have direct expropriation for some realizations of the shock and regulation for other realizations in the our setting, if the host country is privately informed about the shock. Instead, the host country has to choose either private and state ownership. If private ownership is preferred, the simplest way to achieve this is by awarding firms full compensation for all foregone operating profits under direct expropriation: $T^x = \Pi(k)$.

5.4 Monopoly

We have assumed that firms are small and non-strategic in the sense that every individual firm takes prices as given and also neglects any effect of their investment on the ex post probability of being regulated. These assumptions were made for the sake of analytical convenience. But FDI is sometimes undertaken by firms with significant market power in their output markets. One would expect these firms occasionally to be large enough relative to host countries to let their investment decisions be influenced by how they affect the probability of regulation. Indeed, the early FDI literature discusses how investors could reduce host country governments' incentives to expropriate by choosing more complex production techniques than necessary, or by maintaining vital parts of the production process outside the host country; recall the discussion in Section 1.1.

We demonstrate in Appendix A.2 that Proposition 1 applies much more broadly than our assumed production structure would seem to suggest. For instance, it also holds for the case of monopoly and strategic investment. A remaining question is how such market power affects the optimal threshold for regulation.

Let $X(k)$ denote the monopoly production as defined by

$$\Omega_z(X) + \Omega_{zz}(X)X \equiv C_x(X, k),$$

assuming that the second-order condition $\Pi_{kk} < 0$ is fulfilled, and given the price $P(k) \equiv \Omega_z(X(k))$. Absent investment protection, the host country will regulate for $\theta > \Theta(k)$. The optimal investment is

$$k^M \equiv \arg\max_{k \geq 0} [F(\Theta^N(k))\Pi(k) - R(k)].$$

The equilibrium threshold for regulation will hence be $\theta^M \equiv \Theta(k^M)$.

The following Proposition, which we prove in Appendix A.7, identifies circumstances under which the globally efficient outcome (θ^G, k^G) can obtained through an investment agreement:

Proposition 10 *Assume that the foreign investor has monopoly power, that there would be equilibrium overregulation absent any investment agreement ($\theta^M \leq \theta^G$), and that the marginal production externality is in the range*

$$\int_{\underline{\theta}}^{\theta^G} \Psi_k(k^G, \theta) dF(\theta) \in [0, (1 - F(\theta^G))R_k(k^G)].$$

The global welfare optimum (θ^G, k^G) can then be implemented as a sub-game perfect equilibrium by an investment agreement that stipulates the compensation rule

$$T(k, \theta) = \begin{cases} \Pi(k) & \text{if } \theta \leq \theta^M \\ 0 & \text{if } \theta > \theta^M, \end{cases}$$

where $\theta^M = F^{-1}(R_k(k^G)/\Pi_k(k^G)) \geq \theta^G$.

The conditions that render the globally optimal solution feasible under a carve-out policy are qualitatively similar to those under perfect competition established in Proposition 5: first, there is room for a carve-out policy to improve regulatory performance ($\theta^M \leq \theta^G$); second, there is underinvestment because of a non-negative marginal production externality; third, the externality is not so strong as to render optimal full compensation in all states of the world. We conclude that the properties of optimal investment agreements do not depend critically upon market structure and the assumption about non-strategic investors.

5.5 Asymmetric information concerning regulatory shocks

We have thus far assumed ex post verifiability of the economic consequences of regulatory shocks. One can use market data to estimate the market value of an investment and sometimes estimate domestic welfare effects by means of econometric techniques. It is also possible that an observed shock is so severe as to dominate all other effects and render regulation globally optimal. In other cases, it is reasonable to assume that host countries in particular are better informed than outsiders about the domestic welfare effects of regulatory shocks. Certain aspects of investment agreements are better understood within the framework of such host country private information. For instance, in Section 5.3 we pointed to asymmetric information as a rationale for investment agreements to have stricter compensation rules for direct than regulatory expropriation. We will here formally consider the design of agreements when regulatory preferences cannot be directly observed, thus allowing policy makers to misrepresent the true motives of their regulations. Will this fundamentally change the optimal design of investment agreements?

To shed some light on what *can* be achieved when government preferences are not directly observable to outsiders, we will assume that θ is observed only by the government of the host country. It is then no longer possible to implement an agreement that awards compensation below a threshold $\hat{\theta}$ but nothing above. For realizations $\theta > \theta^N$ the host country would simply claim that $\theta > \hat{\theta}$, in order to be allowed to regulate without compensation. Instead, incentive compatibility implies that host countries generally must compensate firms for regulatory intervention (the proof is provided in Appendix A.8):

Proposition 11 *Assume that the host country is privately informed about the shock θ, and that compensation cannot exceed industry profit $\Pi(k)$. Any optimal investment agreement can then be characterized in terms of a threshold $\hat{\theta}$ and a compensation rule*

$$T(k,\theta) = \begin{cases} \Pi(k) & \text{if } \theta \leq \hat{\theta} \\ \max\{-S(k,\hat{\theta}); 0\} & \text{if } \theta > \hat{\theta}. \end{cases}$$

The host country will allow production in country i if $\hat{\theta} \leq \Theta(k)$ and $\theta \leq \Theta(k)$, or if $\hat{\theta} > \Theta(k)$ and $\theta \leq \min\{\hat{\theta}; \Theta^G(k)\}$, and it will regulate otherwise.

The optimal compensation scheme under asymmetric information has the same structure as the one under full information by relying on a threshold value $\hat{\theta}$ for the realized regulatory shocks. They are also identical by requesting full compensation $\Pi(k)$ for all realizations below the threshold value. But there are also important differences that appear for shocks above the threshold $\theta > \hat{\theta}$. Under full information, the host country will not be requested to compensate in case of regulation. But compensation is required in the asymmetric information case. As a result, there will be compensation payments in equilibrium.

The optimal compensation scheme in Proposition 11 violates our constraints on feasible agreements in two closely related respects for $\theta > \hat{\theta}$. First, compensation is not based on investors' foregone operating profit, but on the value to the host country of shutting down production. Second, investors will not receive full compensation for $\theta \in (\hat{\theta}, \Theta^G(k))$, since $-S(k,\hat{\theta}) < \Pi(k)$.[34] In the next section we explore in more detail the efficiency gains that can be achieved by increasing the degree of flexibility in the compensation schemes.

[34] The observation that the compensation is independent of the shock mirrors a standard result in auction theory that the payment is independent of the winner's (unobservable) valuation in an optimal auction (Myerson, 1981).

5.6 Allowing other forms of compensation schemes

Restricting compensation to a simple carve-out policy, i.e. full compensation for foregone operating profit below a threshold and no compensation otherwise, leaves the host country with only the regulatory threshold at its disposal for correcting firms' investment incentives and the ex post incentive to overregulate. This is sufficient if the marginal externality is non-negative, but not too large; see Propositions 5 and 10 and the related discussion. However, when the marginal externality is negative, it is necessary to introduce features that are typically not found in actual agreements, in order to achieve full efficiency. In particular, we move beyond restrictions (1)-(5) specified in Section 2. In what follows, we first review a number of such schemes in the literature to identify how they deviate from the compensation schemes we have considered so far, and we then present an alternative efficient scheme. For reasons of comparison, we recast all models within the context of our current framework.

Hermalin (1995) considers distortions to investments and regulation in a model with direct expropriation and one single firm. He derives two efficient mechanisms. In the first mechanism, a firm pays a production tax equal to the country's value of seizing the asset. In our setting, a tax equal to $-S(k,\theta)$ would implement the globally efficient solution. A tax system sophisticated enough to induce each firm to internalize the full social cost of its actions would render any regulation superfluous: The firm would voluntarily shut down production whenever the social cost exceeded the benefit. The second mechanism instead requests the host country to pay the firm the same amount in compensation subsequent to expropriation. This highlights a fundamental property of efficient compensation, namely that it should be based also on the social cost and not only on operating profit. Even so, Hermalin's (1995) second compensation rule would not be efficient in our setting. It amounts to an expected compensation of $-\int_{\theta^G}^{\bar{\theta}} S(k,\theta)dF(\theta)$, which generally differs from the expected compensation $-\int_{\underline{\theta}}^{\theta^G} S(k,\theta)dF(\theta)$ that generates globally optimal investment incentives. The difference is that the social cost of the investment is based upon production taking place in Hermalin's (1995) analysis, while it is based on being shut down in our setting.

Blume et al (1984) and Aisbett et al (2010a) discuss another form of deviation from the set of feasible compensation schemes that we consider: a carve-out policy with compensation that is linear in the operating profit and the investment cost: $T(k) = \delta\Pi(k) + \alpha R(k)$.[35] This compensation mechanism has two instruments (δ, α) that can be used to correct the distortions to investment and regulation within our framework. It is easy to verify that a special case of this compensation

[35] Compensation schemes based upon other factors that operating profits are not without relevance. The case law sometimes interpret the "full market value" of expropriated assets to mean the value if sold to an outside party. This value could be related to incurred investment costs. However, compensation based on linear combinations of foregone operating profits and incurred investment costs, which is an essential feature of the scheme here, does not seem to occur in reality.

scheme, where

$$T(k,\theta) = \begin{cases} \Pi(k) & \text{if } \theta \leq \theta^G \\ \delta\Pi(k) + (1-\delta)\frac{\Pi(k^G)}{R(k^G)}R(k) & \text{if } \theta > \theta^G \end{cases} \tag{17}$$

would implement the globally efficient outcome (k^G, θ^G) as a Nash equilibrium, if and only if

$$\delta \equiv \frac{F(\theta^G)\Pi(k^G) - \frac{F(\theta^G)}{1-F(\theta^G)}\frac{R(k^G)}{R_k(k^G)}\int_{\underline{\theta}}^{\theta^G}\Psi_k(k^G,\theta)dF(\theta)}{F(\theta^G)\Pi(k^G) - R(k^G) + \frac{R(k^G)}{R_k(k^G)}\int_{\underline{\theta}}^{\theta^G}\Psi_k(k^G,\theta)dF(\theta)} > 0.$$

The efficiency of carve-out policies depends upon verifiability of the regulatory shock θ; see the discussion in Section 5.5. A virtue of (17) is that implementation of the global optimum does not rely upon the shock θ being publicly observable.[36] However, the compensation scheme deviates from those we have previously considered in several respects. First, there are no carve-outs since compensation is paid for any regulation, in line with the scheme in Proposition 11. Second, the scheme in (17) overcompensates firms for their losses since $T(k) > \Pi(k)$ for some investments $k \neq k^G$.[37] Yet another essential feature of this scheme is that it relies upon an ability to estimate the firm's operating profit $\Pi(k^G)$ and capital cost $R(k^G)$ at the efficient investment level k^G. This is of questionable empirical relevance, and it would presumably make its implementation difficult. Both carve-out policies and (17) are derived under the assumption that all firms are identical if an industry consists of more than one firm. In reality, firms in the same industry often differ significantly in terms of size and profits. In this case, a globally efficient carve-out policy would require a unique threshold for compensation for each firm. The linear compensation rule (17) requires firm-specific δ to ensure that each firm faces correct investment incentives. Tailoring an agreement to firm-level characteristics in this fashion can be done in commercial contracts between host countries and individual investors concerning specific projects, but does not occur in state-to-state treaties.

Stähler (2016) derives a mechanism that can implement the globally efficient solution under asymmetric information about θ without any information about k^G. Also, it does not rely upon symmetry. Adapted to our setting, the compensation

$$T(k) = \frac{\widetilde{T} + \int_0^{\Theta^G(k)} S(k,\theta)dF(\theta)}{1 - F(\Theta^G(k))} \tag{18}$$

induces efficient investment if regulation is ex post efficient, i.e. if the host country applies the regulatory threshold $\Theta^G(k)$.[38] In particular, $T(k)$ only depends upon the actual investment k. Ex

[36] Notice that $T(k^G,\theta) = \Pi(k^G) = -S(k^G,\theta^G)$ for all θ implies that the net benefit $S(k^G,\theta) - S(k^G,\theta^G)$ to the host country of allowing production is positive (negative) if $\theta < (>)\theta^G$ evaluated at the investment level k^G.

[37] Generecially, $T_k(k^G) - \Pi_k(k^G) = (1-\delta)(\frac{\Pi(k^G)}{R(k^G)}R_k(k^G) - \Pi_k(k^G)) \neq 0$. This implies $T(k) - \Pi(k) > T(k^G) - \Pi(k^G) = 0$ for some $k \neq k^G$.

[38] This is true even if the firm behaves strategically. Under the compensation rule (18), the expected profit of the firm equals $\int_{\underline{\theta}}^{\Theta^G(k)}(S(k,\theta) + \Pi(k))dF(\theta) - R(k) + \widetilde{T}$ under ex post optimal regulation, which is identical to the social welfare function up to the constant $\widetilde{T}$. The purpose of $\widetilde{T}$ is only to ensure non-negative compensation.

post efficient regulation is ensured by requiring that the country pays compensation $\Pi(k)$. This scheme differs from those in actual agreements since it requires that the host country payment differs from the compensation received by the firm; Stähler (2016) assumes that an arbitrator enables the parties to break the payment balance in this fashion. The compensation rule is thus a Vickrey-Clarke-Groves type of mechanism.

The compensation mechanisms we have considered so far all have their merits, either in terms of relative simplicity (carve-out policies), incentive compatibility (linear compensation as in (17), or non-reliance on efficient investments (as in (18)). But they also have their shortcomings as descriptions of actual agreements, as we have seen. We show in Appendix A.9 that it is possible to implement the globally efficient outcome while still avoiding most or all of these drawbacks, by letting compensation be based on firms' *relative performance*. This could be relevant for cases where the same regulatory intervention affects multiple firms, so that several firms are potentially eligible for compensation. There are many examples of such situations; for instance, the termination of the renewable energy support schemes by Spain and other countries, or the German shut down of nuclear power after Fukushima.

Proposition 12 *A compensation scheme that is based on relative performance can under certain circumstances implement full efficiency even when this cannot be done with the optimal scheme characterized in Proposition 1.*

The specific efficient compensation scheme we identify, see equation (A.18), differs from (17) by being based entirely on the investments that firms have actually made, instead of a counterfactual of what firms would have earned at the global optimum. Hence, the compensation can be estimated based upon historical information about operating profits and the environmental impact of the firm's investment. Furthermore, the host country never overcompensates the firms. The rule differs from (18) by not relying on third-party participation. Instead, the mechanism breaks the balance of payment between the host country and each individual firm by simultaneously adjusting the compensation to *other* firms in the industry. Because the compensation is based upon the performance of similar firms, each firm is compensated for its operating profit in equilibrium.

No information is required about the shock θ to be able to implement the globally efficient outcome. The compensation mechanism is robust to asymmetric information in two other dimensions as well. First, we have previously assumed that policy makers in the host country ignore the effect on operating profits in the decision whether to regulate. Some of the firms could be domestically or even state-owned, or political preferences could affect the way decision makers value profit, and this could be private information (Aisbett et al, 2010a). Policy makers have incentives to exaggerate the extent to which they account for investor profit as basis for an argument that regulation was in fact globally optimal and therefore should not be considered indirect expropriation. The mechanism does not depend upon host country political incentives being observable; see Appendix A.9. Second, we have previously assumed that firms' operating profits are observable. In reality, productivity differences could render operating profit unobservable even if investments were the same across firms and common knowledge. Firms could then have an incentive to exaggerate the value of continued production to increase their compensation. The compensation rule is independent of the firm's own profit, and no firm therefore has any unilateral incentive to misreport it.[39]

The efficiency of the relative performance mechanism does not depend upon firms being identical. What is important, is that each firm can be placed in a comparison group with other similar firms. However, the mechanism does not work if firms are very dissimilar from one another, in particular if the industry consists of one single firm.

6 Concluding remarks

The number of international investment agreements has increased dramatically since the mid-1980s, and protection of foreign direct investment has become a core issue in the policy debate in developed countries. But the economic literature still sheds very limited light on their appropriate role and design. The purpose of this paper is to contribute to filling this void. To this end, we have analyzed the design and implications of optimal investment agreements when constrained to be of the same form as actual agreements. The model, which builds on a canonical model of an investment hold-up problem, reflects the interplay between the two distortions that investment agreements have to address in the case of indirect expropriations: the externalities from investment that motivate the existence of regulation, and the disregard of the interests of foreign investors in regulation decisions.

We have highlighted a large number of characteristics and implications of optimal agreements. Our analysis employs a highly stylized setting for expositional reasons, but we believe that several results are much more general than this setting suggests:

- Optimal investment protection schemes have a very simple structure, essentially being characterized by a threshold value for the intensity of the regulatory shock that determines a carve-out from the compensation requirements. This considerably simplifies the analysis of optimal agreements;

- Optimal investment agreements will not induce host countries to abstain from regulating in situations where regulation would have increased aggregate welfare; that is, there will be no global regulatory chill;

- The incentives to form investment agreements depend on the ability of host countries to make unilateral commitments with regard to investment protection, as well as on the direction of the potential investment flows between the parties;

[39] Myerson and Satterthwaite (1983) derive an optimal compensation mechanism with asymmetric information on both sides when there is a single firm. Their compensation scheme features payments even if there is no regulation and therefore cannot be applied to our setting. They also consider the case of an arbitrator to break the payment balance.

- There are fundamental differences between the rationale and implications for agreements between developed and developing countries, and agreements between developed countries; and
- Investment agreements between developed economies, such as the proposed TTIP, are likely to benefit foreign investors at the expense of rest of society.

There are many important aspects of investment agreements that we have left out. For instance, we have assumed that agreements are formed if and only if this increases national welfare, but we have not modelled the political process of forming investment agreements. Extending the model to include a process of e.g. lobbying, such as in Grossman and Helpman (1995), could shed further light on the politics of investment agreements.

Second, a very common critique in the policy debate is that investment agreements give foreign *investors* legal standing—the ISDS mechanism. A key issue here is how a system that only allows foreign *governments* to litigate would differ from the ISDS system. It seems intuitively plausible that the ISDS system will give more active enforcement, but we do not have any formal support for this conjecture. A similar problem arises in the framework of trade agreements, where little is known about how governments select trade disputes.

Third, we have disregarded the rationale for including investment protection in trade agreements, as is increasingly the practice. Presumably, this has to do with a complementarity between trade and the investment undertakings. It is tempting to view such complementarity as emanating from e.g. global value chains or reflecting exchanges of concessions in the investment and trade areas, but their precise nature is unclear; see Bown and Horn (2015) for an informal discussion, and Maggi (2016) for an analytical taxonomy of various forms of complementarities in trade agreements.

Fourth, the model implicitly assuming that the alternative to FDI in a production facility is to abstain from such investment. In practice, firms can often choose between FDI and arms-length arrangements with local producers when differences in production costs across countries makes it attractive to locate a part of production abroad. As highlighted in the literature on outsourcing, arms-length contracts are typically incomplete and could be a source of hold-up problems between the firms; see Helpman (2006), and Antràs and Rossi-Hansberg (2009) for literature surveys. Extending our setting to include an arms-length option would point to one advantage of such an arrangement, namely that local firms probably are less likely to be regulated than a vertically integrated foreign firm. However, the formation of investment agreements might make such outsourcing less attractive.

Fifth, we have considered the incentives to form a bilateral investment agreement in isolation from the rest of the world. But those incentives might depend on the extent to which other countries form agreements. For instance, the surge of bilateral investment treaties between developed and developing countries is occasionally said to reflect a race-to-the-top between developing countries competing in investment protection to attract foreign investment.

Finally, we have for the most part focused on agreements fulfilling restrictions (1)-(5) in Section

2. We believe that there is strong institutional support for these assumptions. But it would nevertheless be desirable to endogenize some of these incomplete contracting features, possibly using a contracting cost approach, similar to how Horn et al (2010) endogenize the design of trade agreements.

References

[1] 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.

[2] 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) Article 6.

[3] Bergstrand, Jeffrey H. and Peter Egger (2013). "What determines BITs?" *Journal of International Economics* 90(1), 107-122.

[4] Blume, Lawrence, Daniel L. Rubinfeld and Perry Shapiro (1984). "The Taking of Land: When Should Compensation be Paid?" *Quarterly Journal of Economics* 99(1), February, 71-92.

[5] Bown, Chad P. and Henrik Horn (2015). "Investment Protection in Regional Trade Agreements." (Mimeo.).

[6] Bronfenbrenner, Martin (1955). "The Appeal of Confiscation in Economic Development." *Economic Development and Cultural Change* 3(3), 201-218.

[7] Cole, Harold L. and William B. English (1991). "Expropriation and Direct Investment." *Journal of International Economics*, 30(3-4), 201-227.

[8] Crawford, James (2002). *The International Law Commission's Articles on State Responsibility: Introduction, Text and Commentaries.* Cambridge, UK: Cambridge University Press.

[9] Dixit, Avinash (1988). "Strategic Aspects of Trade Policy." In *Advances in Economic Theory*, Thomas Bewley (ed.), Fifth World Congress. Cambridge, MA: Cambridge University Press.

[10] Dixit, Avinash (2011). "International Trade, Foreign Direct Investment, and Security." *Annual Review of Economics* 3(1), 191-213.

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

[12] Eaton, Jonathan and Mark Gersovitz (1983). "Country Risk: Economic Aspects." In Richard J. Herring (ed.), *Managing International Risk*, New York: Cambridge University Press.

[13] Eaton, Jonathan and Mark Gersovitz (1984). "A Theory of Expropriation and Deviations from Perfect Capital Mobility." *The Economic Journal* 94(373), 16-40.

[14] Fudenberg, Drew and Jean Tirole (1991). *Game Theory.* Cambridge, MA: MIT Press.

[15] Hajzler, Christopher (2012). "Expropriation of Foreign Direct Investment: Sectoral Patterns from 1993 to 2006." *Review of World Economics* 148(1), 119-149.

[16] Horn, Henrik (2006). "National Treatment in the GATT." *American Economic Review* 96(1), 394-404.

[17] Horn, Henrik, Giovanni Maggi and Robert W. Staiger (2010). "Trade Agreements as Endogenously Incomplete Contracts." *American Economic Review* 100(1), 394-419.

[18] International Law Commission (2001). "Responsibility of States for Internationally Wrongful Acts." *Yearbook of the International Law Commission*, vol. II (Part Two).

[19] Janeba, Eckhard (2016). "Regulatory Chill and the Effects of Investor State Dispute Settlement." Mimeo, August 3.

[20] Keynes, John M. (1924). "Foreign Investment and the National Advantage." *The Nation and Athenaeum*, August 9, 584-587.

[21] Kohler, Wilhelm and Frank Stähler (2016). "The Economics of Investor Protection: ISDS versus National Treatment." Mimeo, April 22.

[22] Konrad, Kai A. (2016). "Large Investors and Permissive Regulation: Why Environmentalists May Dislike Investor-State Dispute Settlement." Mimeo, September 2.

[23] Lejour, Arjan and Maria Salfi (2015). "The Regional Impact of Bilateral Investment Treaties on Foreign Direct Investment." CPB Discussion Paper 298, CPB Netherlands Bureau for Economic Policy Analysis.

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

[25] Miceli, Thomas J. and Kathleen Segerson (1994). "Regulatory Takings: When Should Compensation Be Paid?" *Journal of Legal Studies* 23(2), June, 749-776.

[26] Miceli, Thomas J. and Kathleen Segerson (2011). "Regulatory Takings. Department of Economics Working Papers Series 2011-16, University of Connecticut.

[27] Myerson, Roger B. (1981) "Optimal Auction Design." *Mathematics of Operations Research* 6(1), 58-73.

[28] Myerson, Roger B. and Mark A. Satterthwaite (1983). "Efficient Mechanisms for Bilateral Trading." *Journal of Economic Theory* 29(2), April, 265-281.

[29] Raff, Horst. 1992. "A Model of Expropriation with Asymmetric Information." *Journal of International Economics* (33)3–4, 245–265.

[30] Schjelderup, Guttorm and Frank Stähler (2016). "Investor State Dispute Settlement and Multinational Firm Behavior." Mimeo, November 3.

[31] 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.

[32] Stähler, Frank (2016). "An Optimal Investor State Dispute Settlement Mechanism." Mimeo, September 2.

[33] Thomas, J. and T. Worral (1994). "Foreign Direct Investment and the Risk of Appropriation." *Review of Economic Studies* 61(1), 81-108.

[34] Turrini, Alessandro and Dieter Urban (2000). "For Whom is MAI? A Theoretical Perspective on Multilateral Agreements on Investment. Mimeo.

[35] Turrini, Alessandro and Dieter Urban (2008). "A theoretical perspective on multilateral agreements on investment." *Review of International Economics* 16(5), 1023–1043.

[36] UNCTAD (2015). *World Investment Report.*

[37] Van Harten, Gus (2016). "Foreign Investor Protections in the Trans-Pacific Partnership." Canadian Centre for Policy Alternatives, June.

A Appendix

A.1 Some comparative statics results

The expressions in (1) yield

$$X^i_k(k_i) = \frac{-C^j_{xk}(X^i(k_i), k_i)}{C^j_{xx}(X^i_k(k_i), k_i) - \Omega^i_{zz}(X^i_k(k_i))} > 0,\ P^i_k(k_i) = \Omega^i_{zz}(X^i(k_i))X^i_k(k_i) < 0.$$

The positive slope of $K^{iN}(\hat{\theta})$ is seen by differentiating (3):

$$K^{iN}_\theta = \frac{-f^i(\hat{\theta}_i)C^j_k(X^i(\hat{k}_i), \hat{k}_i)}{F^i(\hat{\theta}_i)[C^j_{xk}(X^i(\hat{k}_i), \hat{k}_i)X^i_k(\hat{k}_i) + C^j_{kk}(X^i(\hat{k}_i), \hat{k}_i)] + R^j_{kk}(\hat{k}_i)} > 0.$$

Monotonicity follows from $C^j_k < 0$, the concavity of consumer utility and convexity of the production and investment cost:

$$C^j_{xk}X^i_k + C^j_{kk} = \frac{C^j_{xx}C^j_{kk} - C^j_{xk}C^j_{kx} - \Omega^i_{zz}C^j_{kk}}{C^j_{xx} - \Omega^i_{zz}} > 0.$$

The equilibrium without investment protection yields underinvestment relative to the globally efficient level ($k^N_i < k^G_i$) if and only if the marginal expected production externality is sufficiently large (but not necessarily positive):

$$\int_{\underline{\theta}_i}^{\Theta^{iG}(k^N_i)} \Psi^i_k(k^N_i, \theta_i)dF^i(\theta_i) > -R^j_k(k^N_i)\frac{F^i(\Theta^{iG}(k^N_i)) - F^i(\theta^N_i)}{F^i(\theta^N_i)}.$$

A.2 Properties of optimal compensation schemes

The proof of Proposition 1 itself is simple and just a special case of Proposition A.2 below. However, this appendix analyzes the properties of optimal investment schemes in a broader framework than the one employed in the main text, with the purpose of finding common denominators of such schemes. In particular, we consider general non-linear compensation schemes and we place weak restrictions on market structures, ownership and strategic interaction.

The general method of proof is to show that by an appropriate modification to an original compensation scheme, we can ensure that it is ex post optimal for the host country to regulate if and only if the shocks are severe, but without affecting the incentives to invest. Furthermore, this improved regulatory efficiency benefits the host country without any negative consequences for the industry. The subsequent proofs are constructive by explicitly deriving the modified compensation T^{ni} to each firm n in country i as a convex combination of the firm's operating profit Π^{ni} and the original compensation $\hat{T}^{ni}$. As we discuss in more detail below, T^{ni} then inherits many of the properties of $\hat{T}^{ni}$. In particular, improved regulatory efficiency does not depend upon draconian punishments of the host country. For instance, the modified scheme does not rely on punitive damages if this is not a property of the original scheme ($T^{ni} \leq \Pi^{ni}$ if $\hat{T}^i \leq \Pi^i$).

Theorem A.1 below establishes that any optimal investment scheme can be characterized in therms of a threshold that yields domestic, but never global regulatory chill under very general conditions. We allow firms to influence the ex post decision to regulate through their investment. Hence, we solve the market game for a subgame-perfect equilibrium. Theorem A.2 shows that the threshold result extends also to the case when firms perceive their influence on the threshold for regulation as being exogenous to their own investment. In this case, we solve the market game for a Nash equilibrium. We then consider the consequences of more restrictive compensation mechanisms that limit compensation to be proportional and at most equal to operating profit. We show that the optimal investment scheme is a carve-out policy both in subgame-perfect equilibrium (Proposition A.1) and Nash equilibrium (Proposition A.2).

The model. There are two countries, indexed by $i \neq j = 1, 2$, and an industry with $I \geq 1$ firms, indexed by $n = 1, 2..., I$. Assume that each firm n invests k_{ni} in country i, so that $\mathbf{k}_n = (k_{n1}, k_{n2})$ is the firm's investment portfolio. Let $\mathbf{k}_{-ni} = (k_{1i}, .., k_{n-1,i}, k_{n+1,i}, .., k_{Ii})$ be the investment profile of all firms in country i other than n, and denote by $\mathbf{k}_i = (k_{ni}, \mathbf{k}_{-ni})$ the full portfolio of investments in country i. We assume that firms make their investment decisions simultaneously and independently to maximize unilateral profit, but do not make any assumptions about the nature of strategic interaction in the investment stage nor in the product market. Let $\Pi^{ni}(\mathbf{k}_i) \geq 0$ be the reduced form operating profit of firm n of its facilities in country i, and assume that this profit is independent of whether country j is regulated or not. Obviously, $\Pi^{ni}(\mathbf{k}_i) = 0$ if firm n does not have any facilities in country i. Denote by $R^n(\mathbf{k}_n) \geq 0$ firm n's rental cost of capital, which is strictly positive if either $k_{n1} > 0$ or $k_{n2} > 0$.

Let the reduced form domestic welfare be $S^i(\mathbf{k}_i, \theta_i)$ under production and zero if there is regulation. This domestic welfare depends on domestic investment $\mathbf{k}_i$ and on a country-specific shock θ_i. Let domestic welfare be strictly decreasing in θ_i for all $\mathbf{k}_i \geq \mathbf{0}$ (where a weak inequality means that $k_{ni} > 0$ for at least one firm and a strict inequality means that investments are strictly positive for all firms). Assume that $S^i(\mathbf{k}_i, \theta_i)$ and the profit functions are continuous in $\mathbf{k}_i$.

Assume that the two country shocks (θ_1, θ_2) are continuously distributed on the convex domain $\mathcal{A}_1 \times \mathcal{A}_2 \in \mathbb{R}^2$ with marginal cumulative distribution function $F^i(\theta_i)$ and marginal density $f^i(\theta_i)$ in country i. Firms make their investment decisions before the shock is realized, but the countries may choose to regulate subsequent to observing the shock. Regulation implies that the host country disallows the production of all firms in the industry in the host country. Both countries take the decision to regulate simultaneously and independently. The realization (θ_1, θ_2) is common knowledge when the countries make their decisions (although this does not matter because of linear separability between the two countries). Assume that country i attaches the weight $\gamma_{ni} \in [0, 1]$ to the profit of firm n in its decision whether to regulate. Let $\boldsymbol{\gamma}_i = (\gamma_{1i}, .., \gamma_{ni}, .., \gamma_{Ii})$.

Consider now the ex post optimal choice of country i absent any investment agreement. It is then optimal for country i to maintain production if and only if $\theta_i \leq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, where

$$S^i(\mathbf{k}_i, \Theta^i) + \textstyle\sum_{n=1}^{I} \gamma_{ni} \Pi^{ni}(\mathbf{k}_i) \equiv 0.$$

The decision to regulate is independent of country j's actions even if country i may have an interest in what is going on in country j through foreign ownership.[40] Define the threshold $\Theta^{iG}(\mathbf{k}_i) \equiv \Theta^i(\mathbf{k}_i, \mathbf{1}) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. This is the cut-off below which it is ex post optimal to maintain production in country i if national welfare is defined by the sum of domestic welfare and industry operating profit.

International investment agreement (IIA). This is a vector $\hat{\mathbf{T}}^i = (\hat{T}^{1i}, .., \hat{T}^{ni}, .., \hat{T}^{Ii})$ of compensation rules for each country, where $\hat{T}^{ni} : \mathbb{R}_+^I \times \mathcal{A}_i \rightarrow \mathbb{R}_+$ specifies the compensation from the host country to firm n in case of regulation in country i. Notice that the compensation rule only depends on domestic factors. For example, country i never pays out any compensation for regulation abroad. The compensation to each firm is non-negative by assumption. We also assume that the firm receives no compensation under the IIA if the industry remains unregulated.

The timing of the game is as follows:
1. The two countries jointly commit to an IIA with compensation rules $\hat{\mathbf{T}} = (\hat{\mathbf{T}}^1, \hat{\mathbf{T}}^2)$;
2. Firms decide how much capital $\mathbf{k}$ to invest;
3. The shock (θ_1, θ_2) is realized;
4. Both countries observe (θ_1, θ_2) and unilaterally decide whether to regulate.
(a) If country i does not intervene, then product market competition ensues in that country at least.
(b) If country i regulates, then the agreement pays out compensation according to $\hat{\mathbf{T}}^i$.

A subgame-perfect equilibrium (SPE) of the market game induced by IIA $\hat{\mathbf{T}}$ consists of two parts. First, for any investment profile $\mathbf{k}$, the SPE defines two subsets of shock realizations in each country, the set $M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ of θ_i for which the host country allows production and the complementary set $M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ of $\theta_i \in \mathcal{A}_i$ for which the host country prefers to regulate (disallow production):

$$\begin{aligned} M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) &\equiv \{\theta_i : S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^I \gamma_{ni}\Pi^{ni}(\mathbf{k}_i) \geq -\textstyle\sum_{n=1}^I (1-\gamma_{ni})\hat{T}^{ni}(\mathbf{k}_i, \theta_i)\}, \quad \text{(A.1)} \\ M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^{ir}) &\equiv \{\theta_i \notin M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)\}. \end{aligned}$$

Observe that $M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ and $M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ also depend on $\boldsymbol{\gamma}_i$, but for the most part we subsume $\boldsymbol{\gamma}_i$ for notational simplicity. Second, $\hat{\mathbf{k}}_n = (\hat{k}_{n1}, \hat{k}_{n2})$ represents an equilibrium investment profile under IIA $\hat{\mathbf{T}}$ if for all firms $n = 1, 2..., n$:

$$\begin{aligned} \hat{\mathbf{k}}_n \in \; & \underset{\mathbf{k}_n \in \mathbb{R}_+^2}{\arg\max} \{\textstyle\sum_{i=1,2} [\Pi^{ni}(k_{ni}, \hat{\mathbf{k}}_{-ni}) \int_{M^i(k_{ni}, \hat{\mathbf{k}}_{-ni}, \hat{\mathbf{T}}^i)} dF^i(\theta_i) \\ & + \int_{M^{ir}(k_{ni}, \hat{\mathbf{k}}_{-ni}, \hat{\mathbf{T}}^i)} \hat{T}^{ni}(k_{ni}, \hat{\mathbf{k}}_{-ni}, \theta_i) dF^i(\theta_i)] - R^n(\mathbf{k}_n)\}. \end{aligned} \quad \text{(A.2)}$$

In this expression, $\hat{\mathbf{k}}_{-ni} = (\hat{k}_{1i}, .., \hat{k}_{n-1,i}, \hat{k}_{n+1,i}, .., \hat{k}_{ni})$ is the equilibrium investment profile for all firms except n in country i.

Equilibrium expected profit and host country welfare. Let $\hat{M}^i \equiv M^i(\hat{\mathbf{k}}_i, \hat{\mathbf{T}}^i)$ be the subset of shocks realizations in country i for which the host country allows production in equilibrium, and

[40] We assume that the host country allows production if indifferent.

let $\hat{M}^{ir} \equiv M^{ir}(\hat{\mathbf{k}}_i, \hat{\mathbf{T}}^i)$ be the events with equilibrium regulation. Then

$$\tilde{\Pi}^{ni}(\hat{\mathbf{T}}) \equiv \Pi^{ni}(\hat{\mathbf{k}}_i)\int_{\hat{M}^i} dF^i(\theta_i) + \int_{\hat{M}^{ir}} \hat{T}^{ni}(\hat{\mathbf{k}}_i, \theta_i) dF^i(\theta_i) \tag{A.3}$$

is the equilibrium expected operating profit of firm n in market i and

$$\tilde{\Pi}^{n}(\hat{\mathbf{T}}) \equiv \tilde{\Pi}^{n1}(\hat{\mathbf{T}}) + \tilde{\Pi}^{n2}(\hat{\mathbf{T}}) - R^n(\hat{\mathbf{k}}_n),$$

the total expected profit including capital costs. The equilibrium expected welfare of country i equals

$$\begin{aligned} \tilde{V}^i(\hat{\mathbf{T}}, \boldsymbol{\gamma}_i) \equiv & \int_{\hat{M}^i} (S^i(\hat{\mathbf{k}}_i, \theta_i) + \textstyle\sum_{n=1}^I \gamma_{ni} \Pi^{ni}(\hat{\mathbf{k}}_i)) dF^i(\theta_i) \\ & - \int_{\hat{M}^{ir}} \textstyle\sum_{n=1}^I (1-\gamma_{ni}) \hat{T}^{ni}(\hat{\mathbf{k}}_i, \theta_i) dF^i(\theta_i) \\ & + \textstyle\sum_{n=1}^I \gamma_{ni} [\Pi^{nj}(\hat{\mathbf{k}}_j) \int_{\hat{M}^j} dF^j(\theta_j) \\ & + \int_{\hat{M}^{jr}} \hat{T}^{nj}(\mathbf{k}_j, \theta_j) dF^j(\theta_j)] - \textstyle\sum_{n=1}^I \gamma_{ni} R^n(\hat{\mathbf{k}}_n). \end{aligned}$$

Let $\hat{\theta}_i^G \equiv \Theta^{iG}(\hat{\mathbf{k}}_i)$ be the ex post efficient level of regulation given the equilibrium investment $\hat{\mathbf{k}}_i$, so that $S^i(\hat{\mathbf{k}}_i, \hat{\theta}_i^G) = -\sum_{n=1}^I \Pi^{ni}(\hat{\mathbf{k}}_i)$. Define the expected operating surplus in country i as

$$\tilde{W}^i(\hat{\mathbf{T}}) \equiv \int_{\hat{M}^i} (S^i(\hat{\mathbf{k}}_i, \theta_i) - S^i(\hat{\mathbf{k}}_i, \hat{\theta}_i^G)) dF^i(\theta_i). \tag{A.4}$$

We can then write the equilibrium expected welfare of country i more compactly as

$$\tilde{V}^i(\hat{\mathbf{T}}, \boldsymbol{\gamma}_i) = \tilde{W}^i(\hat{\mathbf{T}}) + \textstyle\sum_{n=1}^I [\gamma_{ni} \tilde{\Pi}^n(\hat{\mathbf{T}}) - \tilde{\Pi}^{ni}(\hat{\mathbf{T}})]. \tag{A.5}$$

Theorem A.1 *Any Pareto optimal international investment agreement can be characterized in terms of a threshold $\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i) \in [\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i), \Theta^{iG}(\mathbf{k}_i)]$ for every country i and country investment profile $\mathbf{k}_i$, with the property that country i allows production if and only if $\theta_i \leq \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$.*

Proof: The method of proof is to show that for any IIA with compensation rule $\hat{\mathbf{T}}$ satisfying the appropriate restrictions, there exists another IIA with compensation rule $\mathbf{T}$ satisfying the same restrictions, with the characteristics in the theorem and that is weakly better for both countries than the original agreement.

Defining the alternative investment agreement. We first partition $\mathcal{A}_i$ into a finer subset of shocks for some threshold value $\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ that we will shortly define:

$$\begin{aligned} A^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) &\equiv \{\theta_i \in M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cap (-\infty, \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)]\} \\ A^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) &\equiv \{\theta_i \in M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cap (-\infty, \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)]\} \\ B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) &\equiv \{\theta_i \in M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cap (\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i), \infty)\} \\ B^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) &\equiv \{\theta_i \in M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cap (\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i), \infty)\} \end{aligned}$$

Hence, "A^i" denotes sets of $\theta_i \leq \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, and "$B^i$" sets of $\theta_i > \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. The presence or absence of superscript "r" indicates whether or not there is regulation under the initial agreement $\hat{\mathbf{T}}$. By construction, $A^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cup B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) = M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ and $A^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cup B^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) = M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)$.

The agreement $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$ is characterized by a threshold $\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ for each country given by

$$F^i(\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) \equiv \min\{\textstyle\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i); F^i(\Theta^{iG}(\mathbf{k}_i))\} \tag{A.6}$$

and compensation requirements $\mathbf{T}^i = (T^{1i}, .., T^{Ii})$, where

$$T^{ni}(\mathbf{k}_i, \theta_i) = \begin{cases} \Pi^{ni}(\mathbf{k}_i) & \theta_i \in A^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cup A^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) \\ \tilde{T}^{ni}(\mathbf{k}_i) & \theta_i \in B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) \\ \hat{T}^{ni}(\mathbf{k}_i, \theta_i) & \theta_i \in B^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) \end{cases}, \tag{A.7}$$

and where

$$\begin{aligned} \tilde{T}^{ni}(\mathbf{k}_i) \equiv & \frac{1}{\int_{B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i)} \{\textstyle\int_{A^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)} \hat{T}^{ni}(\mathbf{k}_i, \tilde{\theta}_i) dF^i(\tilde{\theta}_i) \\ & + \max[\textstyle\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) - F^i(\Theta^{iG}(\mathbf{k}_i)); 0] \Pi^{ni}(\mathbf{k}_i)\} \end{aligned} \tag{A.8}$$

if $\int_{B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) > 0$. The reason why Θ^{iH} depends on $\boldsymbol{\gamma}_i$ is because $M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ depends on $\boldsymbol{\gamma}_i$.

Establishing $\Theta^{iH}(\mathbf{k}, \boldsymbol{\gamma}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. If $F^i(\Theta^{iG}(\mathbf{k}_i)) \leq \int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i)$, then $F^i(\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) = F^i(\Theta^{iG}(\mathbf{k}_i)) \geq F^i(\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i))$, where the inequality follows from $\Theta^{iG}(\mathbf{k}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. Assume next that $\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i) \leq F^i(\Theta^{iG}(\mathbf{k}_i)$. The assumption that $S^i(\mathbf{k}_i, \theta_i)$ is strictly decreasing in θ_i and $\hat{T}^{ni}(\mathbf{k}_i, \theta_i) \geq 0$ jointly imply that

$$\begin{aligned} & S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^I (\gamma_{ni} \Pi^{ni}(\mathbf{k}_i) + (1 - \gamma_{ni}) \hat{T}^{ni}(\mathbf{k}_i, \theta_i)) \\ = & S^i(\mathbf{k}_i, \theta_i) - S^i(\mathbf{k}_i, \Theta^i) + \textstyle\sum_{n=1}^I (1 - \gamma_{ni}) \hat{T}^{ni}(\mathbf{k}_i, \theta_i) \end{aligned}$$

is positive for all $\theta_i < \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ in the original agreement. Hence, $(-\infty, \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)) \subset M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ and therefore $F^i(\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) = \int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i) \geq F^i(\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i))$.

It is ex post optimal for country i to allow production under agreement $\mathbf{T}$ iff $\theta_i \leq \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. Consider the incentives for the host country to regulate the industry under an arbitrary investment profile $\mathbf{k}_i$ for agreement $\mathbf{T}$ and for different realizations of the shock θ_i:

(i) $\theta_i \in A^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cup A^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) = (-\infty, \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)]$. By construction of the agreement, the net benefit of allowing production is non-negative for all $\theta_i \leq \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i) \leq \Theta^{iG}(\mathbf{k}_i)$ because in this case

$$S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^I (\gamma_{ni} \Pi^{ni}(\mathbf{k}_i) + (1 - \gamma_{ni}) T^{ni}(\mathbf{k}_i, \theta_i)) = S^i(\mathbf{k}_i, \theta_i) + \sum_{n=1}^I \Pi^{ni}(\mathbf{k}_i) \geq 0,$$

where the we have used $\sum_{n=1}^I \Pi^{ni}(\mathbf{k}_i) = -S^i(\mathbf{k}_i, \Theta^{iG}(\mathbf{k}_i))$ to obtain the weak inequality.

(ii) $\theta_i \in B^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)$. It is optimal to regulate because the compensation function remains the same as before, and it was optimal to regulate already under the initial agreement.

(iii) $\theta_i \in B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ and $\int_{B^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}^i) > 0$. By the construction of $\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$:

$$\int_{B^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) \equiv \int_{A^{ir}(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) + \max\{\int_{M^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) - F(\Theta^{iG}(\mathbf{k}_i)); 0\}. \tag{A.9}$$

Use $\tilde{T}^{ni}(\mathbf{k}_i)$ defined in (A.8) and (A.9) to decompose the net benefit of allowing production in country i as follows:

$$\begin{aligned} & S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^n (\gamma_{ni}\Pi^{ni}(\mathbf{k}_i) + (1-\gamma_{ni})\tilde{T}^{ni}(\mathbf{k}_i)) \\ = \; & \frac{\int_{A^{ir}(\mathbf{k}_i,\hat{\mathbf{T}}^i)} [S^i(\mathbf{k}_i, \theta_i) - S^i(\mathbf{k}_i, \tilde{\theta}_i)] dF^i(\tilde{\theta}_i)}{\int_{B^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i)} \\ & + \frac{\int_{A^{ir}(\mathbf{k}_i,\hat{\mathbf{T}}^i)} [S^i(\mathbf{k}_i, \tilde{\theta}_i) + \sum_{n=1}^n (\gamma_{ni}\Pi^{ni}(\mathbf{k}_i) + (1-\gamma_{ni})\hat{T}^{ni}(\mathbf{k}_i, \tilde{\theta}_i))] dF^i(\tilde{\theta}_i)}{\int_{B^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i)} \\ & + \frac{[S^i(\mathbf{k}_i, \theta_i) - S^i(\mathbf{k}_i, \Theta^{iG}(\mathbf{k}_i))] \max\{\int_{M^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) - F^i(\Theta^{iG}(\mathbf{k}_i)); 0\}}{\int_{B^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i)} \end{aligned}$$

Assume first that $\int_{A^{ir}(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\theta_i) > 0$. In this case, the term on the second row is strictly negative because $S^i_\theta < 0$ and $\theta_i > \Theta^{iH}(\mathbf{k}_i) \geq \tilde{\theta}_i$ for all $\theta_i \in B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ and $\tilde{\theta}_i \in A^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)$. The term on the third row is strictly negative because regulation is optimal under contract $\hat{\mathbf{T}}$ for all $\tilde{\theta}_i \in A^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)$. The term on the third row is zero if $\int_{M^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) \leq F^i(\Theta^{iG}(\mathbf{k}_i))$ and strictly negative otherwise because then $\theta_i > \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i) = \Theta^{iG}(\mathbf{k}_i)$ for all $\theta_i \in B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$. Both terms on the second and third row vanish if $\int_{A^{ir}(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\theta_i) = 0$. But then $\int_{M^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) > F(\Theta^{iG}(\mathbf{k}_i))$ by (A.9) so the third term is strictly negative.

We conclude that it is ex post optimal for the host country to allow production if and only if $\theta_i \leq \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ under the compensation rule $\mathbf{T}$.

Equilibrium investment will be the same under both agreements. By way of the threshold $\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ for regulation defined in (A.6) and the compensation rules (A.7)-(A.8), the expected operating profit of firm n active in country i under the modified agreement $\mathbf{T}$ becomes

$$\begin{aligned} \Pi^{ni}(\mathbf{k}_i) F^i(\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) + \tilde{T}^{ni}(\mathbf{k}_i) \int_{B^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\theta_i) + \int_{B^{ir}(\mathbf{k}_i,\hat{\mathbf{T}}^i)} \hat{T}^{ni}(\mathbf{k}_i, \theta_i) dF^i(\theta_i) \\ = \Pi^{ni}(\mathbf{k}_i) \int_{M^i(\mathbf{k}_i,\hat{\mathbf{T}}^i)} dF^i(\theta_i) + \int_{M^{ir}(\mathbf{k}_i,\hat{\mathbf{T}}^i)} \hat{T}^{ni}(\mathbf{k}_i, \theta_i) dF^i(\theta_i) \end{aligned}$$

after simplifications. This is *exactly* the same expected operating profit as under the original agreement $\hat{\mathbf{T}}$ for every possible investment profile $\mathbf{k}_i$. Hence, $\hat{\mathbf{k}}$ can be sustained as an equilibrium investment profile also under the modified agreement $\mathbf{T}$.

The equilibrium expected operating profits and investment costs are the same under both agreements. It follows directly from the observation that operating profits and the equilibrium investments are the same under both agreements that $\tilde{\Pi}^{n1}(\mathbf{T}) = \tilde{\Pi}^{n1}(\hat{\mathbf{T}})$, $\tilde{\Pi}^{n2}(\mathbf{T}) = \tilde{\Pi}^{n2}(\hat{\mathbf{T}})$ and $\tilde{\Pi}^{n}(\mathbf{T}) = \tilde{\Pi}^{n}(\hat{\mathbf{T}})$ for all n.

The equilibrium expected welfare of both countries is weakly higher under agreement T. The equilibrium welfare of country i equals

$$\begin{aligned}\tilde{V}^i(\mathbf{T},\boldsymbol{\gamma}_i) &\equiv \tilde{W}^i(\mathbf{T})+\textstyle\sum_{n=1}^{I}[\gamma_{ni}\tilde{\Pi}^n(\mathbf{T})-\tilde{\Pi}^{ni}(\mathbf{T})] \\ &= \tilde{W}^i(\mathbf{T})+\textstyle\sum_{n=1}^{I}[\gamma_{ni}\tilde{\Pi}^n(\hat{\mathbf{T}})-\tilde{\Pi}^{ni}(\hat{\mathbf{T}})]\end{aligned}$$

under agreement $\mathbf{T}$, where

$$\tilde{W}^i(\mathbf{T}) \equiv \textstyle\int_{-\infty}^{\hat{\theta}_i^H}(S^i(\hat{\mathbf{k}}_i,\theta_i)-S^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^G))dF^i(\theta_i),$$

$\hat{\theta}_i^H = \Theta^{iH}(\hat{\mathbf{k}}_i,\boldsymbol{\gamma}_i)$, and the second row of $\tilde{V}^i(\mathbf{T},\boldsymbol{\gamma}_i)$ follows from equilibrium profits being the same for all firms under both agreements. Hence,

$$\begin{aligned}\tilde{V}^i(\mathbf{T},\boldsymbol{\gamma}_i)-\tilde{V}^i(\hat{\mathbf{T}},\boldsymbol{\gamma}_i) &= \tilde{W}^i(\mathbf{T})-\tilde{W}^i(\hat{\mathbf{T}}) \\ &= \textstyle\int_{-\infty}^{\hat{\theta}_i^H}(S^i(\hat{\mathbf{k}}_i,\theta_i)-S(\hat{\mathbf{k}}_i,\hat{\theta}_i^G))dF^i(\theta_i) \\ &\quad -\textstyle\int_{\hat{A}^i}(S^i(\hat{\mathbf{k}}_i,\theta_i)-S^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^G))dF^i(\theta_i) \\ &\quad -\textstyle\int_{\hat{B}^i}(S^i(\hat{\mathbf{k}}_i,\theta_i)-S^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^G))dF^i(\theta_i),\end{aligned}$$

where $\hat{A}^i = A^i(\hat{\mathbf{k}}_i,\hat{\mathbf{T}}^i)$ and $\hat{B}^i = B^i(\hat{\mathbf{k}}_i,\hat{\mathbf{T}}^i)$. Adding and subtracting $S^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^H)$ underneath the three integrals yields

$$\begin{aligned}\tilde{W}^i(\mathbf{T})-\tilde{W}^i(\hat{\mathbf{T}}) &= \textstyle\int_{\hat{A}^{ir}}(S^i(\hat{\mathbf{k}}_i,\theta_i)-S^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^H))dF^i(\theta_i) \\ &\quad +\textstyle\int_{\hat{B}^i}(S^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^H)-S^i(\hat{\mathbf{k}}_i,\theta_i))dF^i(\theta_i) \\ &\quad +(S^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^H)-S(\hat{\mathbf{k}}_i,\hat{\theta}_i^G))(F^i(\hat{\theta}_i^H)-\textstyle\int_{\hat{M}^i}dF^i(\theta_i))\end{aligned}$$

after simplifications, where $\hat{A}^{ir} = A^{ir}(\hat{\mathbf{k}}_i,\hat{\mathbf{T}}^i)$. The expressions on the first two rows are both non-negative because S^i is decreasing in θ_i, $\theta_i \le \hat{\theta}_i^H$ in the domain $\hat{A}^{ir}$, and $\theta_i > \hat{\theta}_i^H$ in the domain $\hat{B}^i$. The term on the final row is zero if $\int_{\hat{M}^i}dF^i(\theta_i) \ge F^i(\hat{\theta}_i^G)$ because then $\hat{\theta}_i^H = \hat{\theta}_i^G$. It is zero also if $\int_{\hat{M}^i}dF^i(\theta_i) < F^i(\hat{\theta}_i^G)$ because then $F^i(\hat{\theta}_i^H) = \int_{\hat{M}^i}dF^i(\theta_i)$. It follows that $\tilde{W}^i(\mathbf{T}) \ge \tilde{W}^i(\hat{\mathbf{T}})$ and therefore $\tilde{V}^i(\mathbf{T},\boldsymbol{\gamma}_i) \ge \tilde{V}^i(\hat{\mathbf{T}},\boldsymbol{\gamma}_i)$ for both countries $i = 1, 2$.■

Our method of proof was to show that for any IIA with arbitrary non-negative compensation $\hat{\mathbf{T}}$ that is paid out if and only if the host country disallows production, we can find another compensation rule $\mathbf{T}$ that is paid out if and only if the host country disallows production, that increases regulatory efficiency, but without affecting equilibrium investments. We characterized one such compensation rule, see equations (A.7)-(A.8) in the proof of Theorem A.1, although many other compensation rules can sustain the same result.

Our specific compensation rule yields a compensation T^{ni} to firm n in country i that is a convex combination of that firm's operating profit Π^{ni} and the compensation $\hat{T}^{ni}$ in the original scheme, where the weights on the two components are country-specific and depend on θ_i, but are the same for all firms that have invested in country i. This structure implies that the modified scheme $\mathbf{T}$ inherits a number of characteristics from the original scheme $\hat{\mathbf{T}}$. First, compensation is non-negative because operating profit is non-negative and the original compensation is non-negative ($\Pi^{ni} \geq 0$ and $\hat{T}^{ni} \geq 0$ imply $T^{ni} \geq 0$). Second, it does not rely on excessive compensation (punitive damages) if the original scheme does not ($\hat{T}^{ni} \leq \Pi^{ni}$ implies $T^{ni} \leq \Pi^{ni}$). Third, the modified scheme is non-discriminatory if the original scheme is non-discriminatory. Fourth, the modified compensation rule is linear in operating profit and capital cost if the original scheme has those characteristics. This final property is relevant to the discussion of the concept of fair market value in international investment agreements. One possible interpretation is that all compensation should be based on the realized capital cost $\hat{T}^{ni} = R^{ni}$ for every firm. However, Theorem A.1 shows that linear compensation rules that incorporate both operating profit and capital costs are Pareto superior.

Nash equilibrium (NE). The above results are based upon the assumption that firms rationally foresee and account for the effect of their investment on the probability of being regulated. In this case, SPE is the appropriate equilibrium concept. In other circumstances, it is more relevant to assume that each firm treats the probability of host country intervention as being exogenous to the own investment. For instance, each firm could be so small that it perceives its impact to be negligible to policy makers compared to the industry as a whole, or because the host country has committed to a threshold for intervention that is independent of investment. Then, NE is the appropriate equilibrium concept. Given the investment agreement $\hat{\mathbf{T}}$, an NE defines two subsets of shock realizations in each country, the set $\hat{M}^i$ of θ_i for which the host country allows production and the complementary set $\hat{M}^{ir}$ of θ_i for which the host country prefers to regulate and a vector of equilibrium investments $\hat{\mathbf{k}}_i$ such that regulation is ex post optimal given $\hat{\mathbf{k}}_i$

$$\begin{aligned} \hat{M}^i &\equiv \{\theta_i : S^i(\hat{\mathbf{k}}_i, \theta_i) + \textstyle\sum_{n=1}^{I} \gamma_{ni} \Pi^{ni}(\hat{\mathbf{k}}_i) \geq -\sum_{n=1}^{I}(1-\gamma_{ni})\hat{T}^{ni}(\hat{\mathbf{k}}_i, \theta_i)\}, \\ \hat{M}^{ir} &\equiv \{\theta_i \notin \hat{M}^i\}, \end{aligned} \tag{A.10}$$

and $\hat{\mathbf{k}}_n = (\hat{k}_{n1}, \hat{k}_{n2})$ represents the profit maximizing investment portfolio given $\hat{\mathbf{k}}_{-ni}$ and the assumption that each firm n treats $\hat{M}^i$ and $\hat{M}^{ir}$ as exogenous to $\mathbf{k}_n$:

$$\begin{aligned} \hat{\mathbf{k}}_n \in \; & \underset{\mathbf{k}_n \in \mathbb{R}_+^2}{\arg\max} \{ \textstyle\sum_{i=1,2} [\Pi^{ni}(k_{ni}, \hat{\mathbf{k}}_{-ni}) \int_{\hat{M}^i} dF^i(\theta_i) \\ + \; & \textstyle\int_{\hat{M}^{ir}} \hat{T}^{ni}(k_{ni}, \hat{\mathbf{k}}_{-ni}, \theta_i) dF^i(\theta_i)] - R^n(\mathbf{k}_n)\}. \end{aligned} \tag{A.11}$$

Every SPE is contained in the set of NEs, so $(\hat{M}^i, \hat{M}^{ir})$ and $\hat{\mathbf{k}}_i$ as defined in (A.10)-(A.11) represent Nash equilibrium outcomes of the market game induced by the IIA $\hat{\mathbf{T}}$. The equilibrium expected welfare $\tilde{V}^i(\hat{\mathbf{T}}, \boldsymbol{\gamma}_i)$ of country i and the operating profits $\tilde{\Pi}^{ni}(\hat{\mathbf{T}})$ and $\tilde{\Pi}^n(\hat{\mathbf{T}})$ of each firm n remain unaffected by the change in equilibrium concept.

Theorem A.2 *Assume that investments and regulatory interventions are Nash equilibrium outcomes. Any Pareto optimal international investment agreement that gives rise to equilibrium investments* $\hat{\mathbf{k}}_i$ *can then be characterized in terms of a threshold* $\hat{\theta}_i^H \in [\Theta^i(\hat{\mathbf{k}}_i, \boldsymbol{\gamma}_i), \Theta^{iG}(\hat{\mathbf{k}}_i)]$ *for every country* i *with the property that country* i *maintains production if and only if* $\theta_i \leq \hat{\theta}_i^H$.

Proof: Consider the properties of an alternative investment agreement $\mathbf{T}$ characterized in terms of a threshold $\hat{\theta}_i^H$ given by

$$F^i(\hat{\theta}_i^H) \equiv \min\{\textstyle\int_{\hat{M}^i} dF^i(\theta); F^i(\hat{\theta}_i^G)\} \geq F^i(\Theta^i(\hat{\mathbf{k}}_i, \boldsymbol{\gamma}_i)),$$

and a vector $\mathbf{T}^i = (T^{1i}, .., T^{ni}, .., T^{Ii})$ of compensation rules, where the compensation to each firm is characterized by

$$T^{ni}(\mathbf{k}_i, \theta_i) \equiv \begin{cases} \Pi^{ni}(\mathbf{k}_i) & \theta_i \leq \hat{\theta}_i^H \\ \tilde{T}^{ni}(\mathbf{k}_i) & \theta_i \in \hat{M}^i \cap (\hat{\theta}_i^H, \infty) \\ \hat{T}^{ni}(\mathbf{k}_i, \theta_i) & \theta_i \in \hat{M}^{ir} \cap (\hat{\theta}_i^H, \infty) \end{cases},$$

and where

$$\begin{aligned} \tilde{T}^{ni}(\mathbf{k}_i) \quad \equiv \quad & \frac{1}{\int_{\hat{M}^i \cap (\hat{\theta}_i^H, \infty)} dF^i(\tilde{\theta}_i)} \{\textstyle\int_{\hat{M}^{ir} \cap (-\infty, \hat{\theta}_i^H]} \hat{T}^{ni}(\mathbf{k}_i, \tilde{\theta}_i) dF^i(\tilde{\theta}_i) \\ & + \max\{\textstyle\int_{\hat{M}^i} dF^i(\tilde{\theta}_i) - F^i(\hat{\theta}_i^G); 0\} \Pi^{ni}(\mathbf{k}_i)\} \end{aligned}$$

if $\int_{\hat{M}^i \cap (\hat{\theta}_i^H, \infty)} dF^i(\tilde{\theta}_i) > 0$. The threshold $\hat{\theta}_i^H = \Theta^{iH}(\hat{\mathbf{k}}_i, \boldsymbol{\gamma}_i)$ and the compensation rule above are just a special cases of the more general one considered in Theorem A.1 evaluated at $\mathbf{k}_i = \hat{\mathbf{k}}_i$. Hence, for investments $\hat{\mathbf{k}}_i$, it is optimal for country i to allow production if $\theta_i \leq \hat{\theta}_i^H$ and to regulate if $\theta_i > \hat{\theta}_i^H$. Furthermore, each firm n that expects to be regulated with probability $F^i(\hat{\theta}_i^H)$ and receive compensation according to the above will optimally invest $\hat{\mathbf{k}}_n$ if all other firms invest $\hat{\mathbf{k}}_{-n1}$ and $\hat{\mathbf{k}}_{-n2}$. Hence, the threshold $\hat{\theta}_i^H$ and investment level $\hat{\mathbf{k}}_i$ can be implemented as a Nash Equilibrium by means of the compensation rules $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$. And since the solution concept neither affects the equilibrium investments, the probability of intervention nor the compensations, it follows that the alternative agreement represents a Pareto improvement even under Nash implementation.■

Theorems A.1 and A.2 are explicit about optimal regulation in general and the optimal compensation for sufficiently small values of the shock $\theta_i \leq \hat{\Theta}^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ and $\theta_i \leq \hat{\theta}^H$. However, the theorems are silent about the optimal compensation for $\theta_i > \hat{\Theta}^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ and $\theta_i > \hat{\theta}^H$ because the modified compensation scheme $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$ is defined relative to some initial and arbitrary compensation scheme $\hat{\mathbf{T}} = (\hat{\mathbf{T}}^1, \hat{\mathbf{T}}^2)$ in this case. To move further, let us place more structure on the permissible compensation schemes in international investment agreements.

Proportional compensation. Assume that the compensation to each firm n operating in country i is limited by and proportional to the operating profit:

$$\hat{T}^{ni}(\mathbf{k}_i, \theta_i) \equiv b_i(\theta_i) \Pi^{ni}(\mathbf{k}_i),\ b_i(\theta_i) \in [0, 1] \tag{A.12}$$

for every investment portfolio $\mathbf{k}_i$ and shock θ_i.

A subgame-perfect equilibrium (SPE). We will be interested in each country's unilateral incentive to optimize investment protection, so assume that country j has an arbitrary compensation mechanism $\hat{\mathbf{T}}^j$ and that only country i is restricted to $\hat{\mathbf{T}}^i$ with proportional compensation as in (A.12). An SPE of the game induced by IIA $\hat{\mathbf{T}}$ still defines a production set $M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ and regulation set $M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ by (A.1), but the equilibrium investment condition changes to

$$\begin{aligned}\hat{\mathbf{k}}_n \in \; & \arg\max_{\mathbf{k}_n \in \mathbb{R}^2_+} \{\Pi^{ni}(k_{ni}, \hat{\mathbf{k}}_{-ni})[\textstyle\int_{M^i(k_{ni},\hat{\mathbf{k}}_{-ni},\hat{\mathbf{T}}^i)} dF^i(\theta_i) + \int_{M^{ir}(k_{ni},\hat{\mathbf{k}}_{-ni},\hat{\mathbf{T}}^i)} b_i(\theta_i) dF^i(\theta_i)] \\ + \; & \Pi^{nj}(k_{nj}, \hat{\mathbf{k}}_{-nj}) \textstyle\int_{M^j(k_{nj},\hat{\mathbf{k}}_{-nj},\hat{\mathbf{T}}^j)} dF^j(\theta_j) + \int_{M^{jr}(k_{nj},\hat{\mathbf{k}}_{-nj},\hat{\mathbf{T}}^i)} \hat{T}^{nj}(k_{nj}, \hat{\mathbf{k}}_{-nj}, \theta_j) dF^j(\theta_i) - R^n(\mathbf{k}_n)\}.\end{aligned} \tag{A.13}$$

for all $n = 1, 2..., I$. In this case, Theorem A.1 can be tightened considerably:

Proposition A.1 *Under the restriction that compensation to every firm n operating in country i must be proportional in the sense of equation (A.12), it is optimal for country i in terms of domestic welfare to define a threshold $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ for every possible investment level $\mathbf{k}_i$ and offer compensation of the form*

$$T^{ni}(\mathbf{k}_i, \theta_i) \equiv \begin{cases} \Pi^{ni}(\mathbf{k}_i) & \theta_i \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \\ 0 & \theta_i > \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i). \end{cases} \tag{A.14}$$

It is ex post optimal to maintain production in country i under this compensation rule if and only if $\theta_i \leq \min\{\hat{\Theta}^i(\mathbf{k}_i); \Theta^{iG}(\mathbf{k}_i)\}$. Any Pareto optimal international investment agreement can be characterized in terms of such thresholds and compensation rules if proportionality in compensation applies to both countries.

Proof: A compensation rule $\hat{\mathbf{T}}^i$ that limits the compensation to each firm in country i to at most its operating profit implies

$$S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^I [\gamma_{ni} \Pi^{ni}(\mathbf{k}_i) + (1 - \gamma_{ni}) \hat{T}^{ni}(\mathbf{k}_i, \theta_i)] \leq S^i(\mathbf{k}_i, \theta_i) + \sum_{n=1}^I \Pi^{ni}(\mathbf{k}_i),$$

in which case there will be foreclosure for all shocks above $\theta_i > \Theta^{iG}(\mathbf{k}_i)$ under *any* investment protection scheme. Hence, $M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) \subset (-\infty, \Theta^{iG}(\mathbf{k}_i))$ in the notation of Theorem A.1, which in turn implies

$$F^i(\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) = \textstyle\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta) \leq F^i(\Theta^{iG}(\mathbf{k}_i)). \tag{A.15}$$

Defining the alternative compensation scheme. Consider the threshold $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ defined by

$$F^i(\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)) \equiv \textstyle\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i) + \int_{M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)} b(\theta_i) dF^i(\theta_i) \leq 1 \tag{A.16}$$

and the compensation mechanism (A.14).

Establishing $\Theta^{iH}(\mathbf{k}, \boldsymbol{\gamma}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. A comparison of (A.16) and (A.15) yields $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \geq \Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, whereas $\Theta^{iH}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ by the assumption that compensation is non-negative; see the proof of Theorem A.1.

It is ex post optimal for country i to allow production under $\mathbf{T}^i$ iff $\theta_i \leq \min\{\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i); \Theta^{iG}(\mathbf{k}_i)\}$. The net benefit of allowing production given $\mathbf{T}$ equals

$$S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^I [\gamma_{ni}\Pi^{ni}(\mathbf{k}_i) + (1-\gamma_{ni})T^{ni}(\mathbf{k}_i, \theta_i)] = S^i(\mathbf{k}_i, \theta_i) + \sum_{n=1}^I \Pi^{ni}(\mathbf{k}_i)$$

if $\theta_i \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ and is non-negative if $\theta_i \leq \Theta^{iG}(\mathbf{k}_i)$. We have already shown that it is ex post optimal to disallow production for all $\theta_i > \hat{\Theta}^{iG}(\mathbf{k}_i)$ if compensation to each firm is at most $\Pi^{ni}(\mathbf{k}_i)$. If $\Theta^{iG}(\mathbf{k}_i) \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, we are done. Assume therefore that $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) < \Theta^{iG}(\mathbf{k}_i)$ and consider the net benefit of allowing production for $\theta_i \in (\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i), \Theta^{iG}(\mathbf{k}_i)]$:

$$S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^I [\gamma_{ni}\Pi^{ni}(\mathbf{k}_i) + (1-\gamma_{ni})T^{ni}(\mathbf{k}_i, \theta_i)] = S^i(\mathbf{k}_i, \theta_i) + \sum_{n=1}^I \gamma_{ni}\Pi^{ni}(\mathbf{k}_i),$$

which is strictly negative for all $\theta_i > \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. We conclude that it is ex post optimal for the host country to allow production if and only if $\theta_i \leq \min\{\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i); \Theta^{iG}(\mathbf{k}_i)\}$ under the compensation rule $\mathbf{T}^i$ defined in (A.14).

Equilibrium investments are the same under $\mathbf{T}^i$ as $\hat{\mathbf{T}}^i$. All firms in country i are allowed to produce if $\theta_i \leq \min\{\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i); \Theta^{iG}(\mathbf{k}_i)\}$ given $\mathbf{T}^i$. They are regulated, but never receive any compensation if $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \leq \Theta^{iG}(\mathbf{k}_i)$ and $\theta_i > \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. They are regulated and receive full compensation if $\Theta^{iG}(\mathbf{k}_i) < \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ and $\theta_i \in (\Theta^{iG}(\mathbf{k}_i), \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)]$, but no compensation if $\theta_i > \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. Either way, firm n's expected operating profit in country i equals

$$\Pi^{ni}(\mathbf{k}_i)F^i(\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)) = \Pi^{ni}(\mathbf{k}_i)[\textstyle\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i) + \int_{M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)} b(\theta_i) dF^i(\theta_i)],$$

which is exactly the same expected operating profit as under the original agreement $\hat{\mathbf{T}}^i$ for every possible investment profile $\mathbf{k}_i$. Neither $\hat{\mathbf{T}}^j$ nor the incentives to regulate have changed in country j, so $\hat{\mathbf{k}}$ can be sustained as an equilibrium investment profile also under the compensation schemes $(\mathbf{T}^i, \hat{\mathbf{T}}^j)$.

The equilibrium operating profits and investment costs are the same under $\mathbf{T}^i$ as $\hat{\mathbf{T}}^i$. As operating profits and equilibrium investments are independent of whether country i offers $\hat{\mathbf{T}}^i$ or $\mathbf{T}^i$, it follows that $\tilde{\Pi}^{ni}(\mathbf{T}^i, \hat{\mathbf{T}}^j) = \tilde{\Pi}^{ni}(\hat{\mathbf{T}})$, $\tilde{\Pi}^{nj}(\hat{\mathbf{T}}^j, \mathbf{T}^i) = \tilde{\Pi}^{nj}(\hat{\mathbf{T}})$ and $\tilde{\Pi}^n(\mathbf{T}^i, \hat{\mathbf{T}}^j) = \tilde{\Pi}^n(\hat{\mathbf{T}})$ for all n.

The equilibrium welfare of both countries is weakly higher under agreement $\mathbf{T}^i$ than $\hat{\mathbf{T}}^i$. Welfare in country j is not affected by the change from $\hat{\mathbf{T}}^i$ to $\mathbf{T}^i$ in country i as long as the equilibrium investments are unaltered, because there are no environmental spill-over effects between the two countries. Hence, $\tilde{V}^j(\hat{\mathbf{T}}^j, \mathbf{T}^i, \boldsymbol{\gamma}_j) = \tilde{V}^j(\hat{\mathbf{T}}, \boldsymbol{\gamma}_j)$. Expected welfare in country i is still defined by (A.5) under $\hat{\mathbf{T}}$, and by

$$\tilde{V}^i(\mathbf{T}^i, \hat{\mathbf{T}}^j, \boldsymbol{\gamma}_i) \equiv \tilde{W}^i(\mathbf{T}^i, \hat{\mathbf{T}}^j) + \textstyle\sum_{n=1}^I (\gamma_{ni}\Pi^n(\hat{\mathbf{T}}) - \Pi^{ni}(\hat{\mathbf{T}}))$$

under the alternative configuration $(\mathbf{T}^i, \hat{\mathbf{T}}^j)$ of compensation schemes, where

$$\tilde{W}^i(\mathbf{T}^i, \hat{\mathbf{T}}^j) \equiv \int_{-\infty}^{\min\{\hat{\theta}_i;\hat{\theta}_i^G\}} (S^i(\hat{\mathbf{k}}_i, \theta_i) - S^i(\hat{\mathbf{k}}_i, \hat{\theta}_i^G))dF^i(\theta_i)$$

and $\hat{\theta}_i = \hat{\Theta}^i(\hat{\mathbf{k}}_i)$. The welfare difference equals

$$\begin{aligned}\tilde{V}^i(\mathbf{T}^i, \hat{\mathbf{T}}^j, \boldsymbol{\gamma}_i) - \tilde{V}^i(\hat{\mathbf{T}}, \boldsymbol{\gamma}_i) &= \int_{-\infty}^{\min\{\hat{\theta}_i;\hat{\theta}_i^G\}} (S^i(\hat{\mathbf{k}}_i, \theta_i) - S^i(\hat{\mathbf{k}}_i, \hat{\theta}_i^G))dF^i(\theta_i) \\ &\quad - \int_{\hat{M}^i} (S^i(\hat{\mathbf{k}}_i, \theta_i) - S^i(\hat{\mathbf{k}}_i, \hat{\theta}_i^G))dF^i(\theta_i)\end{aligned}$$

Adding and subtracting $S^i(\hat{\mathbf{k}}_i, \min\{\hat{\theta}_i;\hat{\theta}_i^G\})$ underneath the two integrals yields

$$\begin{aligned}\tilde{V}^i(\mathbf{T}^i, \hat{\mathbf{T}}^j, \boldsymbol{\gamma}_i) - \tilde{V}^i(\hat{\mathbf{T}}, \boldsymbol{\gamma}_i) &= \int_{\hat{M}^{ir}\cap(-\infty,\min\{\hat{\theta}_i;\hat{\theta}_i^G\}]} (S^i(\hat{\mathbf{k}}_i, \theta_i) - S^i(\hat{\mathbf{k}}_i, \min\{\hat{\theta}_i;\hat{\theta}_i^G\}))dF^i(\theta_i) \\ &\quad + \int_{\hat{M}^i\cap(\min\{\hat{\theta}_i;\hat{\theta}_i^G\},\hat{\theta}_i^G]} (S^i(\hat{\mathbf{k}}_i, \min\{\hat{\theta}_i;\hat{\theta}_i^G\}) - S^i(\hat{\mathbf{k}}_i, \theta_i))dF^i(\theta_i) \\ &\quad + [S^i(\hat{\mathbf{k}}_i, \min\{\hat{\theta}_i;\hat{\theta}_i^G\}) - S^i(\hat{\mathbf{k}}_i, \hat{\theta}_i^G)][F^i(\min\{\hat{\theta}_i;\hat{\theta}_i^G\}) - \int_{\hat{M}^i} dF^i(\theta_i)]\end{aligned}$$

after simplification. By the assumption that S^i is decreasing in θ_i, $\theta_i \leq \min\{\hat{\theta}_i;\hat{\theta}_i^G\}$ for all $\theta_i \in \hat{M}^{ir} \cap (-\infty, \min\{\hat{\theta}_i;\hat{\theta}_i^G\}]$ and $\theta_i \geq \min\{\hat{\theta}_i;\hat{\theta}_i^G\}$ for all $\theta_i \in \hat{M}^i \cap (\min\{\hat{\theta}_i;\hat{\theta}_i^G\}, \hat{\theta}_i^G]$, it follows that the expressions on the first two rows are non-negative. The term on the third row is non-negative. The first term in square brackets is non-negative by the assumption that S^i is decreasing in θ_i and $\min\{\hat{\theta}_i;\hat{\theta}_i^G\} \leq \hat{\theta}_i^G$. The second term in square brackets is non-negative because

$$F^i(\hat{\theta}_i) - \int_{\hat{M}^i} dF^i(\theta_i) = \int_{\hat{M}^{ir}} b(\theta_i)dF^i(\theta_i) \geq 0$$

by (A.16) and

$$F^i(\hat{\theta}_i^G) \geq F^i(\hat{\theta}_i^H) = \int_{\hat{M}^i} dF^i(\theta)$$

by (A.15). Hence, $V^i(\mathbf{T}^i, \hat{\mathbf{T}}^j, \boldsymbol{\gamma}_i) \geq V^i(\hat{\mathbf{T}}, \boldsymbol{\gamma}_i)$. It follows that a unilateral deviation by country i from $\hat{\mathbf{T}}^i$ to $\mathbf{T}^i$ represents a Pareto improvement for any $\hat{\mathbf{T}}^j$. Analogously, we can show that a unilateral deviation by country j from $\hat{\mathbf{T}}^j$ to $\mathbf{T}^j$ can be achieved in a Pareto improving manner given $\mathbf{T}^i$ if even country j is restricted to proportional compensation mechanisms. Hence, $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$ is a (weakly) better policy than $\hat{\mathbf{T}}$ for all parties under an international investment agreement under the restriction to proportional compensation.■

Nash Equilibrium. Consider finally the consequences of proportional compensation in country i under the assumption that firms treat the probability of regulation as exogenous to the own investment and the game is solved in terms of Nash equilibrium. An NE of the game induced by $\hat{\mathbf{T}}$ defines a production set $\hat{M}^i$ and regulation set $\hat{M}^{ir}$ by (A.10) as a function of the equilibrium investment profile $\hat{\mathbf{k}}_i$, with the new equilibrium investment condition for all $n = 1, 2..., I$:

$$\begin{aligned}\hat{\mathbf{k}}_n \in \; & \arg\max_{\mathbf{k}_n \in \mathbb{R}_+^2} \{\Pi^{ni}(k_{ni}, \hat{\mathbf{k}}_{-ni})[\int_{\hat{M}^i} dF^i(\theta_i) + \int_{\hat{M}^{ir}} b_i(\theta_i)dF^i(\theta_i)] \\ + \; & \Pi^{nj}(k_{nj}, \hat{\mathbf{k}}_{-nj})\int_{\hat{M}^j} dF^j(\theta_j) + \int_{\hat{M}^{jr}} \hat{T}^{nj}(k_{nj}, \hat{\mathbf{k}}_{-nj}, \theta_j)dF^j(\theta_i) - R^n(\mathbf{k}_n)\}.\end{aligned}$$

Proposition A.2 *Assume that compensation to every firm n operating in country i must be proportional in the sense of equation (A.12). Let investments and regulatory interventions be Nash equilibrium outcomes. In this case, it is optimal for country i in terms of domestic welfare to define a threshold $\hat{\theta}_i$ and offer compensation rules of the form*

$$T^{ni}(\mathbf{k}_i, \theta_i) = \begin{cases} \Pi^{ni}(\mathbf{k}_i) & \text{if } \theta_i \leq \hat{\theta}_i \\ 0 & \text{if } \theta_i > \hat{\theta}_i. \end{cases} \tag{A.17}$$

It is ex post optimal to allow production in country i if (i) $\hat{\theta}_i > \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ and $\theta_i \leq \min\{\hat{\theta}_i; \Theta^{iG}(\mathbf{k}_i)\}$ or; (ii) $\hat{\theta}_i \leq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ and $\theta_i \leq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, but to regulate (disallow production) otherwise. Any Pareto optimal international investment agreement can be characterized in terms of such thresholds and compensation rules if proportionality in compensation applies to both countries.

Proof: Consider the properties of an alternative compensation rule $\mathbf{T}^i = (T^{1i}, .., T^{ni}, .., T^{Ii})$ in country i characterized in terms of a threshold $\hat{\theta}_i$ given by

$$F^i(\hat{\theta}_i) = \int_{\hat{M}^i} dF^i(\theta_i) + \int_{\hat{M}^{ir}} b_i(\theta_i) dF^i(\theta_i) \leq 1$$

and where the compensation to each firm is characterized by (A.17).

Concerning the ex post incentive to regulate, we already know from the proof of Theorem A.1 that it is optimal for country i to allow production for all $\theta_i \leq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ for any mechanism with non-negative compensation. In the proof of Proposition A.1, we also showed that it is ex post optimal to regulate for all $\theta_i > \hat{\Theta}^{iG}(\mathbf{k}_i)$ for any mechanism that restricts the host payment to at most the industry operating profit. If $\hat{\theta}_i > \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, then it is optimal to allow production for all shocks $\theta_i \leq \min\{\hat{\theta}_i; \hat{\Theta}^{iG}(\mathbf{k}_i)\}$ under the proportional mechanism because then

$$S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^{I} [\gamma_{ni} \Pi^{ni}(\mathbf{k}_i) + (1 - \gamma_{ni}) T^{ni}(\mathbf{k}_i, \theta_i)] = S^i(\mathbf{k}_i, \theta_i) + \sum_{n=1}^{I} \Pi^{ni}(\mathbf{k}_i) \geq 0.$$

If $\hat{\theta}_i < \Theta^{iG}(\mathbf{k}_i)$, then it is optimal to regulate for all shocks $\theta_i > \max\{\hat{\theta}_i; \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)\}$ under the proportional mechanism because then

$$S^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{n=1}^{I} [\gamma_{ni} \Pi^{ni}(\mathbf{k}_i) + (1 - \gamma_{ni}) T^{ni}(\mathbf{k}_i, \theta_i)] = S^i(\mathbf{k}_i, \theta_i) - S^i(\mathbf{k}_i, \Theta^i) < 0.$$

With the anticipated regulation level $\hat{\theta}_i$ and proportional compensation rule $T^{ni}(\mathbf{k}_i, \theta_i)$ in country i, and given $\hat{\mathbf{T}}^j$ in country j, it is still optimal for firm n to invest $\hat{\mathbf{k}}_n$ if the other firms maintain their investments at the same level as before. And since the solution concept neither affects the equilibrium investments, the probability of intervention nor the compensations, it follows that the compensation mechanisms $(\mathbf{T}^i, \hat{\mathbf{T}}^j)$ represents a Pareto improvement over $\hat{\mathbf{T}}$ even in Nash equilibrium. In a similar manner, $\mathbf{T}$ represents a Pareto improvement over $(\mathbf{T}^i, \hat{\mathbf{T}}^j)$ if both countries are restricted to proportional compensation.■

Note that Proposition 1 in the main text is a special case of Proposition A.2 above, with $b_i(\theta_i) \in \{0, 1\}$, one representative firm in each country investing only in FDI, and where $\gamma_{ni} = 0$ for the foreign firm n investing in country i.

A.3 Proof of Lemma 1

Let $\hat{\theta}_i^{\mathcal{B}}$ be the firm's consistent belief about investment protection and $\hat{k}_i^B \equiv K^{iN}(\hat{\theta}_i^{\mathcal{B}})$ its profit maximizing investment subsequent to the announcement of $\hat{\theta}_i \leq \theta_i^N$. The firm will earn its full operating profit if $\theta_i \leq \max\{\hat{\theta}_i; \Theta^{iN}(\hat{k}_i^{\mathcal{B}})\}$ and obtain zero profit otherwise. Hence, the firm's beliefs about investment protection is consistent with ex post optimal regulation only if $\hat{\theta}_i^{\mathcal{B}} \in \{\hat{\theta}_i; \theta_i^N\}$ because $\hat{\theta}_i^{\mathcal{B}} = \Theta^{iN}(\hat{k}_i^{\mathcal{B}})$ if and only if $\hat{\theta}_i^{\mathcal{B}} = \theta_i^N$ by assumption (6). Assume that $\hat{\theta}_i < \theta_i^N$ and suppose $\hat{\theta}_i^{\mathcal{B}} = \hat{\theta}_i$. In this case, it is ex post optimal to maintain production if and only if $\theta_i \leq \Theta^{iN}(K^i(\hat{\theta}_i)) > \hat{\theta}_i = \hat{\theta}_i^{\mathcal{B}}$, which is inconsistent. Hence, the only candidate for consistent beliefs is $\hat{\theta}_i^{\mathcal{B}} = \theta_i^N$ for $\hat{\theta}_i \leq \theta_i^N$. The optimal investment then equals $k_i^N = K^i(\theta_i^N)$, and the ex post optimal regulation occurs at $\Theta^i(K^i(\theta_i^N)) = \theta_i^N$, which verifies consistency in this final case.■

A.4 Proof of Lemma 2

Consider first the properties of θ_{NT}^U. Observe that $\tilde{S}(\hat{\theta}) + \tilde{W}(\hat{\theta}) = 2\tilde{S}(\hat{\theta}) + \tilde{\Pi}(\hat{\theta})$ implies a welfare difference

$$\tilde{S}(\theta^U) + \tilde{W}(\theta^U) - \tilde{S}(\hat{\theta}) - \tilde{W}(\hat{\theta}) = 2(\tilde{S}(\theta^U) - \tilde{S}(\hat{\theta})) + \tilde{\Pi}(\theta^U) - \tilde{\Pi}(\hat{\theta}),$$

which is strictly positive for all $\hat{\theta} < \theta^U$. Hence, $\theta_{NT}^U \geq \theta^U$. Alternatively, $\tilde{S}(\hat{\theta}) + \tilde{W}(\hat{\theta}) = 2\tilde{W}(\hat{\theta}) - \tilde{\Pi}(\hat{\theta})$, which implies a welfare difference

$$\tilde{S}(\theta^W) + \tilde{W}(\theta^W) - \tilde{S}(\hat{\theta}) - \tilde{W}(\hat{\theta}) = 2(\tilde{W}(\theta^W) - \tilde{W}(\hat{\theta})) + \tilde{\Pi}(\hat{\theta}) - \tilde{\Pi}(\theta^W),$$

which is strictly positive for all $\hat{\theta} > \theta^W$. Hence, $\theta_{NT}^U \leq \theta^W$. To establish strict inequalities, assume that $\theta_{NT}^U \in (\theta^N, \bar{\theta})$. It is obviously the case that $\theta_{NT}^U > \theta^U$ if $\theta^U = \theta^N$, but $\theta_{NT}^U > \theta^U$ also if $\theta^U > \theta^N$ because then $\tilde{S}_\theta(\theta^U) + \tilde{W}_\theta(\theta^U) = \tilde{\Pi}_\theta(\theta^U) > 0$. Similarly, $\theta_{NT}^U < \theta^W$ if $\theta^W = \bar{\theta}$, but $\theta_{NT}^U < \theta^W$ also if $\theta^W < \bar{\theta}$ because then $\tilde{S}_\theta(\theta^W) + \tilde{W}_\theta(\theta^W) = -\tilde{\Pi}_\theta(\theta^W) < 0$.

Consider next the properties of θ_{NT}^{NS}: $\tilde{S}(\theta_{NT}^U) + \tilde{W}(\theta_{NT}^U) \geq \tilde{S}(\hat{\theta}) - \tilde{W}(\hat{\theta})$ and $\tilde{\Pi}(\theta_{NT}^U) > \tilde{\Pi}(\hat{\theta})$ for all $\hat{\theta} < \theta_{NT}^U$ imply $\theta_{NT}^{NS} \geq \theta_{NT}^U$. The inequality is strict if $\theta_{NT}^U \in (\theta^N, \bar{\theta})$ because then $\tilde{S}_\theta(\theta_{NT}^U) + \tilde{W}_\theta(\theta_{NT}^U) = 0$, but $\tilde{\Pi}_\theta(\theta_{NT}^U) > 0$.■

A.5 Proof of Proposition 5

The second derivative of the welfare function $\tilde{W}(\hat{\theta}) = \tilde{S}(\hat{\theta}) + \tilde{\Pi}(\hat{\theta})$ equals

$$\tilde{W}_{\theta\theta}(\hat{\theta}) = \tilde{W}_\theta(\hat{\theta}) \frac{K_{\theta\theta}(\hat{\theta})}{K_\theta(\hat{\theta})} + [\frac{d}{dk} \int_{-\infty}^{\Theta^G(\hat{k})} [S_k(\hat{k}, \theta) + \Pi_k(\hat{k})] dF(\theta) - R_{kk}(\hat{k})](K_\theta(\hat{\theta}))^2$$

for $\hat{\theta} \geq \theta^E$. Every solution $\tilde{W}_\theta(\hat{\theta}) = 0$ in the domain $\hat{\theta} \geq \theta^E$ is a local maximum by the concavity assumption (13). Hence, $\tilde{W}(\hat{\theta})$ is strictly quasi-concave in the domain $\hat{\theta} \geq \theta^E$.

Part (a): The marginal expected welfare is strictly negative for all $\hat{\theta} \geq \theta^E$ if (14) is satisfied: $\tilde{W}_\theta(\hat{\theta}) \leq \tilde{W}_\theta(\theta^E) < 0$. Hence, $\theta^{NN} < \theta^E$. We already know from Lemma 1 that $\theta^{NN} > \theta^N$. By the stability condition (6), it follows that $\theta^{NN} \in (\Theta(k^{NN}), \Theta^G(k^{NN}))$.

Part (b): The marginal expected welfare satisfies $\tilde{W}_\theta(\theta^E) \geq 0$ and $\tilde{W}_\theta(\bar{\theta}) \leq 0$ if (14) is violated, but (16) is satisfied. In this case, there exists a $\theta^{NN} \in [\theta^E, \bar{\theta}]$ such that $\tilde{W}_\theta(\theta^{NN}) = 0$. As k^G is the unique welfare maximizing investment when regulation is ex post efficient, it follows that $k^{NN} = K(\theta^{NN}) = k^G$. Furthermore $\theta^{NN} \geq \Theta^G(k^{NN}) > \Theta(k^{NN})$ by stability (6) implies that the ex post optimal threshold for regulation is $\min\{\theta^{NN}, \Theta^G(k^{NN})\} = \Theta^G(k^{NN}) = \Theta^G(k^G) = \theta^G$.

Part (c): Strict concavity of the total welfare function and $\tilde{W}_\theta(\bar{\theta}) > 0$ imply that the maximal investment is optimal in the domain $[K(\underline{\theta}), \bar{k}]$. Hence, the optimal level of investment protection is $\theta^{NN} = \bar{\theta}$ in this case. ■

A.6 Proof of Proposition 9

Let θ' be given by $K(\theta') \equiv k^G$, where the function $K(\hat{\theta})$ was defined in Section 3 by the first-order condition (3). Since we are assuming an initial underinvestment, it follows that $\theta' > \Theta(k^G)$. Also, recall $\theta^G = \Theta^G(k^G)$. Consider the following compensation rule, assuming $\theta' < \theta^G$:

$$T(k, \theta) = \begin{cases} \Pi(k) & \text{if } \theta \leq \theta' \text{ or } \theta > \theta^G \text{ and direct expropriation} \\ \Pi(k) & \text{if } \theta \leq \theta^G \text{ and regulation} \\ 0 & \text{if } \theta' < \theta \leq \theta^G \text{ and direct expropriation} \\ 0 & \text{if } \theta > \theta^G \text{ and regulation.} \end{cases}$$

The agreement thus either pays full or no compensation, and it allows the host country to directly expropriate, but not to regulate, without compensation for $\theta \leq \theta' < \theta^G$. Assume that firms have invested k^G. For $\theta < \theta'$ the host country has to pay full compensation both under direct expropriation and regulation. It has no strict incentive to intervene in this case because $\theta' < \theta^G$, which is the critical value beyond which the host country is willing to pay full compensation in order to terminate production for the investment k^G. For $\theta' < \theta \leq \theta^G$, the host country would prefer not to regulate since it then has to pay full compensation. But since it can expropriate directly without compensation, it will do so instead. For $\theta > \theta^G$ it will regulate, and not pay any compensation. Hence, given the investment k^G production will be maintained for $\theta \leq \theta^G$, which is globally efficient.

Now turn to investors. They will not be compensated for host country measures that deprive them of their operating profits for $\theta > \theta'$, but are assured full compensation for any $\theta \leq \theta'$. Hence, k^G fulfills the first–order condition (3).■

A.7 Proof of Proposition 10

Let κ be the set of k satisfying $\theta^M \geq \Theta^N(k)$. Observe that $k^G \in \kappa$ by assumption because $\Theta^N(k^G) \leq \theta^G \leq \theta^M$. It is ex post optimal for the host country to allow production for investments $k \in \kappa$ if and only if $\theta \leq \min\{\theta^M; \Theta^G(k)\}$. The expected monopoly profit equals $F(\theta^M)\Pi(k) - R(k) = \frac{R_k(k^G)}{\Pi_k(k^G)}\Pi(k) - R(k)$ in this case. Obviously, k^G is the profit-maximizing investment in the domain κ. Let $\pi^G \equiv \frac{R_k(k^G)}{\Pi_k(k^G)}\Pi(k^G) - R(k^G)$. For investments $k \notin \kappa$, it is ex post optimal for the host country to maintain production if and only if $\theta \leq \Theta^N(k)$. Hence, the expected monopoly profit equals $F(\Theta^N(k))\Pi(k) - R(k) \leq F(\theta^N)\Pi(k^N) - R(k^N) \equiv \pi^N$ for all $k \notin \kappa$. It follows that k^G is even a globally profit-maximizing investment level because

$$\pi^G - \pi^N = [F(\theta^M)\Pi(k^G) - R(k^G) - F(\theta^M)\Pi(k^N) + R(k^N)] + (F(\theta^M) - F(\theta^N))\Pi(k^N) \geq 0.$$

The term in square brackets is non-negative because k^G maximizes $F(\theta^M)\Pi(k) - R(k)$. The second term is non-negative by the assumption that $\theta^M \geq \theta^G \geq \theta^N$. Given the equilibrium investment level k^G, it is ex post optimal for the host country to maintain production if and only if $\theta \leq \min\{\theta^M; \theta^G\} = \theta^G$.■

A.8 Proof of Proposition 11

Assume that compensation is paid out only if the firm is regulated and that compensation is not allowed to be higher than $\Pi(k)$. Assume also that the representative firm in the host country treats the probability of regulation as exogenous to the own investment k. By the Revelation Principle, we can restrict attention to direct compensation mechanisms (the host country reports θ) that are incentive compatible (the host country cannot benefit from lying about θ). A general compensation mechanism within this framework specifies a probability $\xi(\theta)$ that production is allowed and a compensation $\hat{T}(k,\theta)$ that is paid out in case the firm is regulated.

The equilibrium rent of the host country is

$$V(k,\theta) \equiv \xi(\theta)S(k,\theta) - (1 - \xi(\theta))\hat{T}(k,\theta).$$

By standard arguments (e.g. Fudenberg and Tirole, 1991), the compensation scheme is incentive compatible only if $V_\theta(k,\theta) = \xi(\theta)S_\theta(k,\theta_i)$ and $\xi(\theta)$ is non-increasing in θ. Integrating up yields the expected rent

$$V(k,\theta) = \int_{\underline{\theta}}^{\theta} \xi(\tilde{\theta})S_\theta(k,\tilde{\theta})d\tilde{\theta} + V(k,\underline{\theta}).$$

We can then solve for the incentive compatible compensation by subtracting the two expressions for $V(k,\theta)$ from one another:

$$(1 - \xi(\theta))\hat{T}(k,\theta) = \xi(\theta)S(k,\theta) - \int_{\underline{\theta}}^{\theta} \xi(\tilde{\theta})S_\theta(k,\tilde{\theta})d\tilde{\theta} - V(k,\underline{\theta}).$$

To make the problem economically interesting, assume that it is strictly better to allow production than to regulate for the most favorable shock $\underline{\theta}$, so that $V(k,\underline{\theta}) = S(k,\underline{\theta})$. Assume also that the mechanism does not randomize between production and regulation. Non-randomization and the restriction that $\xi(\theta)$ is non-increasing in θ imply a threshold $\hat{\theta} > \underline{\theta}$ such that $\xi(\theta) = 1$ if $\theta \leq \hat{\theta}$ and $\xi(\theta) = 0$ if $\theta > \hat{\theta}$. We have restricted $\hat{T}(k,\theta)$ to be zero for $\theta \leq \hat{\theta}$. If $\theta > \hat{\theta}$, then

$$\hat{T}(k,\theta) = -\int_{\underline{\theta}}^{\hat{\theta}} S_\theta(k,\tilde{\theta})d\tilde{\theta} - V(k,\underline{\theta}) = -S(k,\hat{\theta}).$$

It is impossible to implement a threshold $\hat{\theta} < \Theta^N(k)$ because this would imply negative compensation:

$$\hat{T}(k,\theta) = -S(k,\hat{\theta}) < -S(k,\Theta^N(k)) = 0 \text{ for all } \theta \in (\hat{\theta},\Theta^N(k)).$$

It is also impossible to implement a threshold $\hat{\theta} > \Theta^G(k)$ because doing so would require overcompensating the firm,

$$\hat{T}(k,\theta) = -S(k,\hat{\theta}) > -S(k,\Theta^G(k)) = \Pi(k) \text{ for all } \theta > \hat{\theta},$$

which we have ruled out by assumption.

Let $\bar{\Theta}(k) \equiv \Theta^N(k)$ if $\hat{\theta} \leq \Theta^N(k)$ and $\bar{\Theta}(k) \equiv \min\{\hat{\theta};\Theta^G(k)\}$ if $\hat{\theta} > \Theta^N(k)$. Then

$$\hat{T}(k,\theta) = \begin{cases} 0 & \text{if } \theta \leq \bar{\Theta}(k) \\ -S(k,\bar{\Theta}(k)) & \text{if } \theta > \bar{\Theta}(k) \end{cases}$$

represents the optimal payment to firms under asymmetric information about the shock $\hat{\theta}$.

A straightforward way to implement the cut-off $\bar{\Theta}(k)$ and payment $\hat{T}(k,\theta)$ would be to decentralize the choice of regulation to the host country and require it to pay the fixed compensation $-S(k,\bar{\Theta}(k))$ whenever it disallows production. In this case, the net benefit $S(k,\theta) - S(k,\bar{\Theta}(k))$ of allowing production would be non-negative if and only if $\theta \leq \bar{\Theta}(k)$.

An alternative compensation rule that emphasizes the role of asymmetric information compared to the optimal compensation scheme under complete information in Proposition 1, would still be to decentralize the decision to regulate to the country, but require it to report θ and pay compensation

$$T(k,\theta) = \begin{cases} \Pi(k) & \text{if } \theta \leq \hat{\theta} \\ \max\{-S(k,\hat{\theta});0\} & \text{if } \theta > \hat{\theta} \end{cases}$$

depending on its report.

To see that this compensation scheme yields the same outcome as above, assume first that $\hat{\theta} \leq \Theta^N(k)$. The country would always report $\theta > \hat{\theta}$ subsequent to regulation in order to pay zero compensation: $\max\{-S(k,\hat{\theta});0\} = 0$ because $S(k,\hat{\theta}) \geq S(k,\Theta^N(k)) = 0$. As the host country would never have to pay compensation for regulation, it would allow production for all $\theta \leq \Theta^N(k)$ and regulate for all $\theta > \Theta^N(k)$.

In the second case, $\hat{\theta} \in (\Theta^N(k),\Theta^G(k)]$, the host country would report $\widetilde{\theta} > \hat{\theta}$ subsequent to regulation because doing so would minimize the compensation payment: $-S(k,\hat{\theta}) \leq \Pi(k)$. In this case, the net benefit $S(k,\theta) - S(k,\hat{\theta})$ of allowing production would be non-negative if and only if $\theta \leq \hat{\theta}$. If $\theta > \hat{\theta}$, then the host country would regulate, truthfully report θ and pay compensation $-S(k,\hat{\theta}) > 0$.

In the third case, $\Theta^G(k) < \hat{\theta}$, the host country would minimize the compensation payment subsequent to regulation by reporting $\tilde{\theta} \leq \hat{\theta}$ because $\Pi(k) < -S(k,\hat{\theta})$ in this case. The net benefit $S(k,\theta)+\Pi(k)$ of allowing production would be non-negative if and only if $\theta \leq \Theta^G(k)$. If $\theta > \Theta^G(k)$, then the host country would regulate, but perhaps misreport $\widetilde{\theta} \neq \hat{\theta}$ to reduce the compensation payment to $\Pi(k)$.■

A.9 A compensation scheme based upon relative performance

Assume that the industry in the host country consists of $I \geq 2$ symmetric foreign firms—the results hold also for some degree of firm heterogeneity. We index firms by $n \neq \hat{n} = 1,...,I$. Let k_n be the investment of firm n and $\mathbf{k} = (k_n, \mathbf{k}_{-n})$ the investment profile of all firms, where $\mathbf{k}_{-n} = (k_1,...,k_{n-1},k_{n+1},...k_I)$ represents the investment profile of all firms other than n. We can then write demand, price, consumer surplus and so forth as functions of $\mathbf{k}$. In particular, the operating profit of firm n is $\Pi^n(\mathbf{k}) \equiv \hat{\Pi}(P(\mathbf{k}), k_n)$. We maintain the assumption that firms are price-takers, so that $-C_k(X(k_n, \mathbf{k}_{-n}), k_n)$ represents the marginal perceived effect of increasing investment k_n on firm n's operating profit, where $X(k_n, \mathbf{k}_{-n})$ is the supply of firm n. Because of perfect competition, each firm treats the operating profit of the other firms in the industry as constant and independent of its own investment.

The threshold $\Theta^G(\mathbf{k})$ for ex post optimal regulation is implicitly defined by

$$S(\mathbf{k}, \Theta^G) + \textstyle\sum_{n=1}^{I} \Pi^n(\mathbf{k}) \equiv 0.$$

The globally optimal investment profile $\mathbf{k}^G$ features the same investment k^G by all firms because of symmetry, and the globally optimal threshold for regulation is $\theta^G = \Theta^G(\mathbf{k}^G)$.

Let

$$\Delta\tilde{\Psi}^n(\mathbf{k}) \equiv \int_{-\infty}^{\Theta^G(\mathbf{k})} (\Psi(\mathbf{k},\theta) - \Psi(0, \mathbf{k}_{-n}, \theta))dF(\theta)$$

be the expected externality associated with firm n's investment if regulation is ex post efficient. Assume that $\Psi_{k_n k_n} \geq 0$ for all n and that each firm n treats all other firms' externality as exogenous to the own investment k_n.[41]

Consider now a relative compensation scheme. A subset $\mathcal{I}(\mathbf{k})$ of all firms form a comparison group of size $|\mathcal{I}(\mathbf{k})|$. Let $\mathcal{I}(\mathbf{k})$ be the largest-sized comparison group such that the compensation scheme

$$T^n(\mathbf{k}) = \begin{cases} \frac{1}{|\mathcal{I}(\mathbf{k})|-1} \sum_{\hat{n}\in\mathcal{I}(\mathbf{k})\backslash n} [\Pi^{\hat{n}}(\mathbf{k}) + \frac{\Delta\tilde{\Psi}^n(\mathbf{k}) - \Delta\tilde{\Psi}^{\hat{n}}(\mathbf{k})}{1-F(\Theta^G(\mathbf{k}))}] & \forall n \in \mathcal{I}(\mathbf{k}) \\ 0 & \forall n \notin \mathcal{I}(\mathbf{k}) \end{cases} \tag{A.18}$$

yields non-negative compensation for all firms in the industry. Here, the compensation depends not only on operating profit, but also on the externality. For instance, the firm receives a relatively large compensation if the externality of its investment is positive compared to that of the other firms in the industry.[42]

[41] Independence is a behavioral assumption here, but could be affected by technology. If the externality is additive, $\Psi(\mathbf{k},\theta) = \sum_{n=1}^{I} \hat{\Psi}^n(k_n,\theta)$, then $\Delta\tilde{\Psi}^{\hat{n}}(\mathbf{k}) = \int_{-\infty}^{\hat{\theta}} [\hat{\Psi}^{\hat{n}}(k_{\hat{n}},\theta) - \hat{\Psi}^{\hat{n}}(0,\theta)]dF(\theta)$, which is independent of k_n for all $\hat{n} \neq n$ if firm n also treats the probability $\hat{\theta}$ of regulation as exogenous to its own investment k_n.

[42] The compensation rule (A.18) is defined only for $|\mathcal{I}(\mathbf{k})| \geq 2$. For completeness, assume that the firm with the maximal $\Pi^n(\mathbf{k}) + \frac{\Delta\tilde{\Psi}^n(\mathbf{k})}{1-F(\Theta^G(\mathbf{k}))}$ is compensated by $\Pi^n(\mathbf{k})$ and that the rest of the firms receive nothing in compensation if $|\mathcal{I}(\mathbf{k})| = 1$.

The total payment is balanced and never implies overcompensating the industry by the host country for any possible investment profile $\mathbf{k}$ or realization of the shock θ:

$$\sum_{n=1}^{I} T^n(\mathbf{k}) = \sum_{n\in\mathcal{I}(\mathbf{k})} \Pi^n(\mathbf{k}) \leq \sum_{n=1}^{I} \Pi^n(\mathbf{k}).$$

In particular, the comparison group contains the entire industry ($|\mathcal{I}(\mathbf{k})| = I$) if the firms have chosen similar investment levels. In this case, the host country must pay the total industry profit in compensation and therefore has an ex post optimal incentive to regulate.

The expected profit of firm n is

$$F(\Theta^G(\mathbf{k}))\Pi^n(\mathbf{k}) + (1 - F(\Theta^G(\mathbf{k})))T^n(\mathbf{k}) - R(k_n)$$

under ex post optimal regulation. Holding the threshold fixed at θ^G, and assuming $\mathbf{k}_{-n} = \mathbf{k}^G_{-n}$, the perceived marginal effect

$$-F(\theta^G)C_k(X(k_n, \mathbf{k}^G_{-n}), k_n) - R_k(k_n) + \int_{-\infty}^{\theta^G} \Psi_{k_n}(k_n, \mathbf{k}^G_{-n}, \theta)dF(\theta)$$

on the expected profit of increasing investment k_n is exactly the same as the marginal expected welfare effect. By the construction of the mechanism (A.18), the host country and the firms all internalize the full economic effects of their actions.

Proposition A.3 *Assume that there are $I \geq 2$ identical foreign firms in the industry and that all firms treat prices, the probability of regulation and the environmental impact of the other firms as exogenous to the own investment. Assume also that the operating profit at the global optimum is sufficiently large: $\Pi(\mathbf{k}^G) - R(k^G) \geq \max_{k\geq 0}\{F(\theta^G)\hat{\Pi}(P(\mathbf{k}^G), k) - R(k)\}$. In this case, the global welfare optimum $(\mathbf{k}^G, \theta^G)$ can be implemented as a Nash equilibrium by an international investment agreement stipulating relative compensation according to (A.18).*

Proof: The host country must pay the full industry profit in compensation if $\mathbf{k} = \mathbf{k}^G$ and will therefore allow production if and only if $\theta \leq \theta^G$. Assume that $\mathbf{k}_{-n} = \mathbf{k}^G_{-n}$ and consider the choice of k_n under the assumption that firm n expects to be regulated with probability θ^G. By strict concavity of the profit function, $k_n = k^G$ is the profit maximizing investment in the domain $k_n \leq \kappa$, where $\kappa > k^G$ is the upper bound to firm n's investment that yields a strictly positive compensation under regulation. The expected equilibrium profit $\Pi(\mathbf{k}^G) - R(k^G)$ by assumption is larger than the maximum profit, $\max_{k\geq 0}\{F(\theta^G)\hat{\Pi}(P(\mathbf{k}^G), k) - R(k)\}$, the firm could obtain if it received no compensation. This is also a necessary condition for implementation of the globally efficient solution under asymmetric information. Hence, k^G is firm n's profit maximizing investment for all $k_n \geq 0$. By continuity, the proposition holds also for some degree firm heterogeneity.■

Implementation of the efficient outcome is independent of any information concerning the extent to which the host country internalizes operating profit. Assume that the host country attaches a weight $\gamma_n \in [0,1]$ to the operating profit of firm n. In this case, the net benefit of allowing production equals $S(\mathbf{k}^G,\theta)+\sum_{n=1}^{I}[\gamma_n \Pi^n(\mathbf{k}^G)+(1-\gamma_n)T^n(\mathbf{k}^G)]$ at the globally efficient investment profile $\mathbf{k}^G$. This is equal to $S(\mathbf{k}^G,\theta)+\sum_{n=1}^{I}\Pi^n(\mathbf{k}^G)$ under (A.18) and therefore independent of all $\gamma_n$because $T^n(\mathbf{k}^G)=\Pi^n(\mathbf{k}^G)$ for all n.

Proposition A.3 holds for some degree of firm heterogeneity. The important part is that firms are sufficiently similar that $|\mathcal{I}(\mathbf{k}^G)|=I$, so that the ex post incentive to regulate is efficient at $\mathbf{k}^G$. The mechanism is still efficient with large firm differences. This happens if the industry can be partitioned into a multiple comparison groups with two or more similar firms in each group such that all of them receive positive compensation in a neighborhood around $\mathbf{k}^G$.