ECONOMICS OF INTERNATIONAL INVESTMENT AGREEMENTS[1]

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

Thomas Tangerås[3]
The Research Institute of Industrial Economics (IFN), Stockholm

First version: November 18, 2016
This version: May 21, 2018

[1]We are grateful for helpful discussions with Robert L. Howse and Petros C. Mavroidis, and for comments by participants during presentations at the Research Institute for Industrial Economics, Stockholm, at the conference "Globalization and New Technology: Effects on Firms and Workers" at Waxholm, at the 1st Applecross Workshop in Economics, and at the conference "Creative Perspectives on International Trade and Foreign Direct Investment: A Celebration to Honor Jim Markusen" at Boulder. The authors also gratefully acknowledge financial support from Riksbankens Jubileumsfond (Horn), the Marianne and Marcus Wallenberg Foundation (Horn), and Jan Wallanders och Tom Hedelius Stiftelse (Tangerås).
[2]Address: Research Institute of Industrial Economics (IFN); Box 55665; SE-102 15 Stockholm; Sweden. Email: henrik.horn@ifn.se.
[3]Address: Research Institute of Industrial Economics (IFN); Box 55665; SE-102 15 Stockholm; Sweden. Email: thomas.tangeras@ifn.se.

Abstract

Close to 2 700 state-to-state investment agreements (IIAs) worldwide protect foreign direct investment (FDI) against host country policies. We analyze the design and implications of protection against regulatory expropriations in IIAs, emphasizing the role of externalities from FDI, countries' unilateral commitment possibilities, and the direction of investment flows. We show e.g. that (*i*) simple compensation mechanisms found in IIAs have desirable efficiency properties; (*ii*) optimal agreements do not cause underregulation ("regulatory chill"); (*iii*) IIAs can have strong distributional effects by benefitting investors at the expense of the rest of society; and (*iv*) IIAs should go further than only to impose non-discrimination.

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

Contents

1 Introduction

International investment agreements (IIAs) are state-to-state treaties that aim to promote foreign direct investment (FDI) by protecting foreign investors against host country policy measures. IIAs typically require host countries to compensate investors in case of expropriation or for measures with similar effects, and they contain a range of other provisions, including non-discrimination of foreign investment. The agreements also normally include investor-state dispute settlement (ISDS) mechanisms that enable foreign investors to litigate against host countries outside host country legal systems. There are approximately 2 700 IIAs currently in force. Most of these agreements are bilateral investment treaties (BITs) that solely address investment protection. But since the formation of NAFTA it has become increasingly common for preferential trade agreements to encompass such protection; for instance, this is a standard feature of newer EU and US preferential trade agreements.[1]

IIAs were initially formed without much political opposition, but they have recently become intensively debated.[2] A number of developing countries have reduced the ambit of their agreements, or terminated them. There has also been a debate in developed countries concerning investment protection in e.g. NAFTA, the Trans-Pacific Partnership (TPP), the Canada-EU Comprehensive Economic and Trade Agreement (CETA), and the EU-US Transatlantic Trade and Investment Partnership (TTIP). These debates have been partly motivated by a number of recent high-profile IIA investment disputes.[3] The critique has concerned both the substantive undertakings and the dispute settlement mechanisms in the IIAs. A main contention is that these agreements cause host countries to abstain from desirable regulations—that is, that they cause "regulatory chill." For instance, U.S. Trade Representative Robert Lighthizer recently stated :

> "...More importantly, we had situations where real regulation which should be in place, which is bipartisan and in everybody's interest, has not been put in place for fears of ISDS...."[4]

The widespread discontent with IIAs raises a number of important questions regarding their design and impact, in particular in light of the enormous investment stocks and flows that they cover. But the economic literature hardly provides any answers to these question. Apart from

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

[2] See e.g. Howse (2017) and Stiglitz (2008) for comprehensive critical discussions of investment agreements.

[3] Contentious cases include the threat by TransCanada Corporation to litigate against the US under NAFTA regarding US$ 15 billion in damages for the Obama administration's decision (overturned by the Trump administration) to disallow the construction of the Keystone XL pipe line; Phillip Morris' litigation against several countries over tobacco plain packaging legislation; litigations against Spain and other countries for withdrawals of renewable energy support schemes; and energy company Vattenfall's litigation against Germany regarding the German decision to accelerate the phase-out of nuclear power in the wake of the Fukushima disaster.

[4] Statement made regarding the rengotiation of NAFTA before the House Ways and Means Committee on March 21, 2018. https://www.c-span.org/video/?c4719932/brady-lighthizer-isds-discussion.

an empirical literature on the effects of IIAs on investment flows, the literature is extremely thin. This contrasts sharply with the voluminous literature on the complementary form of international economic integration—trade agreements.

The purpose of this paper is to contribute to the development of a theory of investment agreements, and in the process illuminate a number of claims in the policy debate. The paper focuses on a main source of contention regarding the substantive undertakings in IIAs: the provisions concerning *regulatory (or indirect) expropriation*.[5] These obligations govern compensation in cases where host country regulatory measures deprive investors of the return on their investments, but where assets are not formally seized. The scope of these provisions is increasingly often restricted by "carve-outs" from the compensation requirements for certain regulatory policies, such as non-discriminatory measures that protect the environment or public health.

In the paper we analyze *optimally designed* agreements with carve-outs. Since the performance of such agreements are likely to be sensitive to the contracting possibilities that the parties have at their disposal, we restrict attention to agreements that *share core features with actual agreements*. For instance, we assume that compensation can only be paid in case of regulation, and that compensation cannot exceed foregone operating profits. This is the first paper to examine investment protection from this angle, to the best of our knowledge.

Turning to the formal setting, an analysis of an agreement regarding regulatory expropriation requires three components: there must be some benefit from FDI for the host country, there must be some reason why there is at least occasionally overregulation, and there must be some rationale for allowing regulation at all (otherwise the optimal agreement would just disallow regulation). Following the seminal contribution by Aisbett et al. (2010a) (see literature review below), we employ a generalized version of the canonical regulatory takings (hold-up) model of investment protection. At the outset, a representative foreign firm makes an irreversible direct investment in a production facility in a host country.[6] Production creates benefits such as employment, higher incomes or technology transfers. But the investment also causes adverse effects, such as environmental damage or health hazards, that can render production undesirable from a domestic and even a joint welfare perspective. The investment can therefore be downward or upward distorted depending on whether additional investment creates positive or negative expected marginal investment externalities in the host country.

The severity of the adverse effects from investment becomes known only after the investment has been sunk. Having observed this "regulatory shock," the host country decides whether to allow production or to shut down the facility (regulate). The host country disregards any loss of operating profits suffered by the foreign investor from a decision to regulate. This disregard of foreign effects generally causes host country overregulation, the expectation of which reduces the foreign investor's

[5] The other main criticized substantive obligation provides for "fair and equitable treatment."

[6] A single investor setting is employed for expositional reasons. Several main results are derived for more general settings in the Appendix.

willingness to invest. The stage is thus set for an investment agreement to correct the distortions to investment and regulation. To capture core features of actual IIAs, we assume among other things that an agreement requires full compensation for foregone operating profits whenever regulation is compensable. But it allows the host country to regulate without compensation if the regulatory shock is sufficiently severe.

Our model generates a large number of findings that we believe yield valuable insights regarding the legal and economic characteristics of investment agreements, and the validity of arguments in the policy debate. We first establish a link between the fundamental legal principles underlying compensation schemes in actual agreements and their efficiency properties. By the simple structure of these schemes, the host and the source country have a single instrument— the level of investment protection—at their disposal to improve efficiency. Nevertheless, an agreement can correct the distortions to both investment and regulation, and thus *implement the unconstrained joint optimum,* under a robust set of circumstances. When this is infeasible, a joint surplus maximizing agreement must strike a balance between correcting overinvestment and overregulation.

We also establish that the compensation scheme has desirable and robust efficiency properties in two fundamental respects. First, for a general class of compensation schemes, any Pareto optimal agreement will implement a *threshold value for regulation*, such that there will be production for regulatory shocks that are weaker than the threshold value, and regulation for more severe shocks. Second, in the class of compensation schemes that are non-decreasing functions of investments, it is joint welfare maximizing to give *full* compensation in case of regulation for regulatory shocks below a threshold value, and *zero* compensation for more severe shocks. This is in line with the structure of compensation payments under actual IIAs.

Second, it is sometimes alleged that compensation payments and disputes under IIAs indicate flaws in the investment regime. Some view disputes as opportunistic exploitation of excessive investment protection by investors. Others instead see the disputes as caused by bad faith regulations by host countries, and thus as proof of the need for investment protection. We offer a third interpretation. In the present setting, equilibrium compensation payments (and possibly also litigations in order to the trigger those payments) are necessary to achieve full efficiency. The payments then serve as *implicit subsidies* to investment, the benefit of which materializes for other realizations of the regulatory shock than those for which the compensation payments are made.

Third, a key policy issue is whether agreements cause regulatory chill, i.e. under-regulation. We provide a general argument for why *Pareto optimal investment agreements never yield under-regulation from a joint welfare perspective.* Agreements instead induce either ex post efficient regulation or inefficient overregulation. They do yield less regulation than would result without any agreement, but this is simply the price the host country pays to promote foreign investment.

Fourth, another much debated issue is the distribution of the benefits and costs of IIAs. We show that this distribution depends both on the direction of investment flows, and on countries' unilateral commitment possibilities regarding investment protection. The purpose of the traditional BIT between a developed country (North) and a developing country (South) is to stimulate investment from the developed to the developing country. The role of the agreement is to enhance the credibility of commitments by South to protect investment from North. As all gains originate in South, South will enter into an agreement with North if and only if it increases domestic welfare in South. In contrast, "North-North" agreements such as CETA or TTIP are meant to stimulate investment flows in both directions. External enforcement of investment protection is less important as the countries' domestic legal systems normally yield sufficiently credible investment protection. Instead, countries have incentives to offer—from a domestic welfare viewpoint excessive—investment protection in return for better protection of their own foreign investments in the partner country. Hence, North-North agreements are likely to *increase foreign investor profits, but reduce domestic welfare in both countries*. These findings are consistent with the considerable popular resistance to agreements such as CETA and TTIP.

The basic model is then extended in a number of directions, to shed some light on a range of issues that have arisen in the debate. First, it is sometimes held that IIAs should go no further than to ensure non-discriminatory treatment of foreign investment. This finds little support in our setting, however: a National Treatment clause alone would have ambiguous welfare consequences compared to the case of no agreement in a North-North setting, and would benefit North and have ambiguous welfare consequences for South, in a North-South setting. Second, most IIAs have stricter rules for direct than for regulatory expropriations, allowing carve-outs from compensation requirements only for the latter. This might seem reasonable, if viewing direct expropriations essentially as unproductive (or worse) transfers of rents. But as **is** [will be] shown, there is an *economic rationale for allowing uncompensated direct expropriation* under certain circumstances.

Third, a contentious policy issue is whether foreign investors should be insured against political risks. A simple reinterpretation of the model that incorporates uncertainty regarding future government preferences, rather than regulatory risks, suggests that ex ante optimal investment agreements formed behind a "veil of political ignorance," should allow governments that are sufficiently sensitive to the regulatory problem to regulate without paying compensation, whereas less sensitive governments should pay compensation. Such an agreement would thus be *sensitive to "democratic" concerns* in this limited regard.

In our final analytical section we discuss how other compensation schemes than those typically found in actual IIAs can achieve efficient outcomes. In particular, we demonstrate how a fully efficient outcome can be implemented using a relative performance scheme that is less informationally demanding than those analyzed in the literature. We also consider asymmetric information more generally, as certain aspects of IIAs can be better understood as resulting from information asymmetries.

The literature Expropriation of FDI has been studied in the economic literature at least since Keynes (1924). But investment treaties received hardly any attention until the late 1990s. The literature instead focused on how implicit mechanisms such as host countries' concerns for their reputations, might alleviate underinvestment problems.[7] A smaller economic literature has since then emerged, with Markusen (1998, 2001) as one of the first contributors. A dominant issue in this literature has been whether IIAs in practice stimulate investment, and there are studies on e.g. the role of a multilateral investment agreement, and the determinants of the formation of IIAs, including the relationship to preferential trade agreements (see e.g. Bergstrand and Egger, 2013).

A nascent literature speaks more directly to the issues raised in the recent policy debate. We will comment on specific issues in this literature below, but it can broadly speaking be divided into two strands. A first strand, represented by Janeba (2016), Kohler and Stähler (2016), Konrad (2016), and Schjelderup and Stähler (2016), considers implications of exogenously specified investment agreements. This approach is useful for considering specific features of IIAs, and in particular the effects of legal provisions that are hard derive from an economic framework.

Our paper is closer to the second strand, which examines the optimal design of investment agreements. Aisbett et al. (2010a) were the first to carefully analyze implications of an optimally designed agreement, building on the regulatory takings literature. While key contributions to the takings literature implicitly assume that the incentives to invest and to regulate are undistorted, Aisbett et al. (2010a) show how an investment agreement can implement full efficiency also with distorted incentives to regulate, even if arbitrators of disputes under the agreement are imperfectly informed regarding the true magnitude of regulatory shocks. Their solution requires the host country to overcompensate the industry for its losses in certain instances. Aisbett et al. (2010b) examine the implications of a National Treatment provision that prevents a host country from requesting up-front payments from foreign firms prior to investing. The provision can render broader exemptions from the compensation requirements desirable in case of regulation. Finally, Stähler (2016) derives a mechanism that can achieve a fully efficient outcome when the regulatory shock is unobservable, assuming that the payment balance between the host country and compensated firms can be broken. Our paper differs from these contributions in several respects. We analyze externalities and their resulting distortions in a more general framework and consider the direction of investment flows as well as the possibilities for countries to make credible unilateral commitments to investment protection through domestic laws and regulations. Most importantly, the three studies above rely on compensation schemes that allow subsidization and overcompensation, or payments to or from third parties. These features are rarely found in actual IIAs, and are assumed away here.

2 Investment and regulation absent an agreement

This Section lays out the model, identifies two fundamental distortions, and characterizes the equilibrium absent investment protection.

[7] Dixit (2011) discusses a range of issues related to insecurity of property rights and FDI, and reviews the literature.

IIAs typically apply to all (or almost all) industries in the countries involved. The optimal design of IIAs can therefore be highly complex due to the interaction between firms within and across industries. We here impose symmetry both within each industry by assuming that firms face identical decision problems, and across industries by assuming that all industries are identical. This allows us to focus on a representative firm in a representative industry, and thus also to avoid firm and industry indexation. But we consider more general settings in the Appendix. We do not make any assumptions regarding market structure, and for expositional convenience normalize the number of firms to $n = 1$, and refer to this as "the firm".

At the outset of the interaction, the representative firm decides how much capital $k \geq 0$ to invest in a production facility in a host country. After the investment has been made, an industry-specific shock θ is realized that affects the net benefit to the host country of allowing production. Having observed the common knowledge shock, the host country unilaterally decides whether to permit production, or to regulate by closing the industry. Ex ante, the shock is continuously distributed on $[\underline{\theta}, \bar{\theta}]$ with cumulative distribution function $F(\theta)$ and density $f(\theta)$. For most of the analysis we will take the representative firm to be small enough relative to the industry for investors to disregard any individual impact on regulatory decisions—firms are "regulation-takers".[8] We consider a "regulation-strategic" monopoly in Section 4.9 and in the Appendix, showing how fully efficient outcomes might be easier to implement in this case.

A firm's investment cost $R(k) \geq 0$ is irreversible and a strictly increasing, weakly convex function of the investment k. The investor receives the operating profit $\pi = \Pi(k) \geq 0$ if the host country allows production, which is strictly increasing and strictly concave in k. The operating profit is zero in case of regulation, this being the sole consequence of regulation for the foreign country.

The investment creates benefits to the host country in terms of consumer surplus, employment, technological spill-overs, learning-by-doing in the work-force, and so forth. But investment can also have adverse consequences, such as pollution or other health hazards. The net benefit $v = V(k, \theta)$ to the host country of allowing production relative to regulating depends on k and the magnitude θ of the shock. The *marginal* net benefit of investment can be positive or negative, $V_k(k, \theta) \gtrless 0$ (subscripts on functional operators denote partial derivatives throughout), but is strictly concave in k.[9] A larger shock θ yields a smaller net value of allowing production ex post of the investment: $V_\theta(k, \theta) < 0$. High realizations of θ could represent the arrival of severely adverse information regarding environmental or health consequences of the production process or the goods produced, or other factors affecting the desirability of the investment. We will not adopt any particular interpretation, but simply denote θ as capturing a "regulatory shock." We assume that any positive and negative externalities arise during the production stage, and that they appear only in case of production. To ensure that there is a role to play for investment and regulation, we assume that

[8] A percevied independence between investment and regulation could also arise if there is long time lag between the investment decision and the regulatory intervention.

[9] We assume that functions $\Pi(k)$, $R(k)$, and $V(k, \theta)$ are twice continuously differentiable.

for any $k > 0$, it is better for the host country to allow production if the shock is sufficiently mild, $V(k,\underline{\theta}) > 0$, and to regulate if it is sufficiently severe, $V(k,\bar{\theta}) < 0$.

The rest of this Section considers the outcome absent an investment agreement, to set the stage for the analysis of an investment agreement in the remainder of the paper. Throughout the analysis we derive the equilibrium outcome in standard fashion by solving for the interaction backwards.

2.1 The regulatory distortion

Absent investment protection, it is optimal for the host country to regulate for shocks that are more severe than $\Theta(k) \in (\underline{\theta}, \bar{\theta})$, as defined by

$$V(k,\Theta) \equiv 0, \tag{1}$$

assuming that the host country allows production if indifferent.

We use the joint surplus $V(k,\theta) + \Pi(k)$ of the host country and the foreign industry as a benchmark for evaluating the efficiency of regulatory intervention; we will briefly discuss the case where the objective function for the regulator differs from national welfare in Section 4.4. The ex post jointly optimal threshold for regulation $\Theta^J(k)$ is given by

$$V(k,\Theta^J) + \Pi(k) \equiv 0.$$

It follows from $\Pi(k) > 0$ and $V_\theta(k,\theta) < 0$ that $\Theta^J(k) > \Theta(k)$ for $k > 0$.

Observation 1 *Absent investment protection and for any investment* $k > 0$, *from a joint surplus perspective the host country:*
(a) correctly allows production for $\theta \leq \Theta(k)$*;*
(b) overregulates for $\theta \in (\Theta(k), \Theta^J(k))$*; and*
(c) correctly regulates for $\theta \geq \Theta^J(k)$*.*

This overregulation for any given investment k is one of the two distortions that an agreement on investment protection might address. We will refer to it as the *regulatory distortion.*

2.2 The investment distortion

A firm that anticipates a threshold value $\hat{\theta}$ for regulation, invests to maximize the expected profit $F(\hat{\theta})\Pi(k) - R(k)$. The profit maximizing investment is an increasing function $K(\hat{\theta})$ of the regulatory threshold $\hat{\theta}$, given by the associated first-order condition

$$F(\hat{\theta})\Pi_k(K) - R_k(K) = 0. \tag{2}$$

The joint expected surplus of the host country and the firm equals

$$\int_{\underline{\theta}}^{\hat{\theta}} [V(k,\theta)dF(\theta) + \Pi(k)]dF(\theta) - R(k) \tag{3}$$

at the regulatory threshold $\hat{\theta}$. The jointly optimal level of investment $K^J(\hat{\theta})$ for any given regulatory threshold, is given by

$$F(\hat{\theta})\Pi_k(K^J) - R_k(K^J) + X(K^J, \hat{\theta}) = 0, \tag{4}$$

where

$$X(k,\hat{\theta}) \equiv \int_{\underline{\theta}}^{\hat{\theta}} V_k(k,\theta)dF(\theta) \tag{5}$$

captures *the expected marginal investment externalities* in the host country evaluated at $(k, \hat{\theta})$. These marginal externalities can be positive or negative as they incorporate both external benefits and costs from FDI for the host country. A comparison of (2) and (4), and using $X_k(k, \hat{\theta}) < 0$, yields the following straightforward result:

Observation 2 *Absent investment protection and for any regulatory threshold $\hat{\theta}$, foreign direct investment is too low from a joint surplus perspective, $K(\hat{\theta}) < K^J(\hat{\theta})$, if and only if $X(K(\hat{\theta}), \hat{\theta}) > 0$.*

This *investment distortion* is the second of the two distortions that an agreement on investment protection needs to address in our model.

2.3 The inefficiency of the outcome

Assume that the foreign firm treats the host country's decision to regulate as exogenous to its own investment. In the Nash equilibrium (k^0, θ^0) of the game between the host country and the firm, the investment decision is then optimal given the equilibrium threshold for regulation, $k^0 = K(\theta^0)$, and the threshold for regulation is ex post optimal given the equilibrium level of investment $\theta^0 = \Theta(k^0)$. To ensure the existence of a unique equilibrium, we assume throughout that

$$\theta' < \Theta(K(\theta')) \text{ iff } \theta' < \theta^0. \tag{6}$$

where θ^0 is given by $\theta^0 = \Theta(K(\theta^0))$.[10] This assumption corresponds to the "stability" conditions used for instance in oligopoly models to rule out counter-intuitive comparative statics properties.[11]

[10] Assumption (6) implies that there exists at most one solution $\theta^0 - \Theta(K(\theta^0)) = 0$. Existence of θ^0 follows by way of the Mean-Value Theorem, $\underline{\theta} - \Theta(K(\underline{\theta})) \leq 0$ and $\bar{\theta} - \Theta(K(\bar{\theta})) \geq 0$. The equilibrium investment $k^0 = K(\theta^0)$ is unique by monotonicity of K.

[11] For instance, this condition is necesssary and sufficient to ensure that the direct reduction in investment that would result from an increase in the marginal cost of investment $R_k(k)$, would not induce the host country to reduce its regulatory interventions to the extent that the equilibrium level of investment increases.

Symmetrically, the jointly optimal solution (k^J, θ^J) is characterized by $k^J = K^J(\theta^J)$ and $\theta^J = \Theta^J(K^J(\theta^J))$, and to ensure uniqueness of (k^J, θ^J) we assume throughout that

$$\theta' < \Theta^J(K^J(\theta')) \text{ iff } \theta' < \theta^J; \tag{7}$$

see the proof of Lemma 1.

The Nash equilibrium (k^0, θ^0) incorporates both an investment and a regulatory distortion, and will therefore generally differ from the jointly optimal solution (k^J, θ^J). But we have not imposed structure to unambiguously determine any qualitative relationship between the two outcomes. The following result is established in Appendix A.1:

Lemma 1 *The Nash equilibrium absent investment protection yields:*
(a) underinvestment relative to the joint optimum ($k^0 < k^J$) if and only if

$$X(k^0, \Theta^J(k^0)) + [F(\Theta^J(k^0)) - F(\theta^0)]\Pi_k(k^0) > 0; \tag{8}$$

(b) overregulation relative to the joint optimum ($\theta^0 < \theta^J$) if and only if

$$V(K^J(\theta^0), \theta^0) + \Pi(K^J(\theta^0)) > 0. \tag{9}$$

Part (a) shows that a sufficient condition for underinvestment is that the expected marginal investment externalities are positive at the Nash equilibrium. This is not a necessary condition however, since the regulatory distortion also tends to cause underinvestment. Similarly, the regulatory distortion does not suffice to ensure overregulation. The resulting tendency toward underinvestment, which will affect the optimal regulation, can more than offset the regulatory distortion: $\theta^0 = \Theta(k^0) < \Theta^J(k^0) \lesseqgtr \Theta^J(k^J) = \theta^J$.[12] According to Part (b), there will be equilibrium overregulation if the joint surplus from allowing production is positive when evaluated at the efficient investment level $K^J(\theta^0)$. Note that the Nash equilibrium can feature overregulation and underinvestment for both positive and negative expected marginal investment externalities.

3 Negotiating an investment agreement

IIAs are meant to stimulate investment by protecting investors against losses due to a range of government actions. The previous Section showed how overregulation and underinvestment can arise in our setting absent investment protection. We will now examine how an investment agreement would affect the outcome. We start by discussing how to formalize salient features of such an agreement.

[12] Expression (8) is derived by evaluating the derivative of the expression for joint expected welfare at k^0.

3.1 The agreement

IIAs are long-term state-to-state commitments to protect foreign investments in a broad range of industries against host country policy interventions.[13] The agreements typically do not include any general commitments regarding subsidization (or taxation) of investment, nor any commitments vis-à-vis individual foreign investors. The rules regarding regulatory expropriations request compensation to be paid only in the case of regulation. Also, according to case law, a government measure must deprive investors of almost all their profits in order to possibly constitute indirect expropriation.

IIAs increasingly include carve-outs from compensation requirements in case of indirect expropriation for actions that protect "legitimate" public welfare objectives, such as public health, safety and the environment. For instance, according to the *US Model Bilateral Investment Treaty 2012,* "[n]on-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."

With regard to the magnitude of compensation, IIAs normally build on basic principles concerning state responsibility in Customary International Law, according to which "...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," and "[t]he compensation shall cover any financially assessable damage including loss of profits insofar as it is established."[14] But IIAs do not allow for punitive payments.[15] Also, any host country payment is received in full by the affected firm(s).

To capture these salient features of actual IIAs, we will focus on agreements with the following features:

(a) The agreements do not contract directly on investment levels or regulation.

(b) Compensation is paid only in case of regulation.

(c) There are no payments to or from outside parties.

(d) Compensation is not required for regulation that protects the host country against sufficiently severe regulatory shocks θ.

(e) Any compensation equals foregone operating profits.

We thus assume that an agreement stipulates for each country a function

$$T = T(\pi, \theta, \theta') \equiv \begin{cases} \pi & \text{if } \theta \leq \theta' \\ 0 & \text{if } \theta > \theta' \end{cases} \tag{10}$$

[13] See Dolzer and Schreuer (2012) for a comprehensive overview of International Investment Law.

[14] The first quote is an often cited passage in the determination by the Permanent Court of International Justice in the case *The Factory at Chorzów* from 1928. The second quote is Article 36, International Law Commission (2001), with a footnote omitted.

[15] According to Crawford (2002, p. 219), "[a] tribunal shall not award punitive damages."

that specifies the compensation the country should pay any regulated investor.[16,17] We denote the cut-off level for compensation θ' as the *level of investment protection* of foreign investment, because firms in the industry are guaranteed their full operating profits for all shocks below θ', but no compensation above. As the same threshold θ' applies to all firms in a particular industry, either all firms receive compensation or no one does; the only firm-specific component in the compensation scheme is hence the requirement that compensation should equal each firm's foregone operating profits. These carve-outs are determined in negotiations between the contracting parties in the analysis to follow.[18]

For most of the analysis we assume that the cut-off value for the carve-out is independent of the level of investment, that is, that θ' in (10) is a scalar rather than some function of investment levels. This is for expositional reasons only. We will show that this assumption does not affect the outcome an agreement, as long as investors do not behave strategically with regard to the impact of their investment on the probability of regulation, as we will mostly assume. But we will allow for the compensation scheme to depend on the investment level in this regard when considering strategic investments in Section 4.9, and in some of the settings examined in the Appendix.

Another characteristic feature of IIAs is their highly potent enforcement mechanisms. IIAs commonly build on *The Convention on the Recognition and Enforcement of Foreign Arbitral Awards* (the New York Convention), which requests courts of the (currently 157) contracting states to recognize and enforce arbitration awards made in any other contracting state. Hence, IIAs can be said to build on compulsory *third-party* enforcement. We therefore assume that the agreements are perfectly enforceable. This is also assumed in much of the trade agreement literature, despite the much weaker formal enforcement mechanisms in those agreements.

3.2 The contracting parties

The parties to actual IIAs differ in a large number of ways, and the agreements are likely to reflect these differences. We will focus on two types of differences between partner countries we believe are fundamental to how agreements function, but appear not to have been highlighted in the literature.

[16] The same specification is employed by Aisbett et al. (2010a).

[17] To define more generally an investment agreement, let each host country c have I_c industries, with H_{ci} foreign-owned firms operating in sector i. Each country is exposed to a vector $(\theta_{c1}, ..., \theta_{ci}, ..., \theta_{cI_c})$ of industry-specific shocks. An investment agreement is then a vector $(T^{c1}, .., T^{ci}, .., T^{cI_c})$ of compensation functions for each partner country c, where $T^{ci} = (T^{ci1}, .., T^{cih}, .., T^{ciH_{ci}})$ specifies the compensation to firm h in industry i in country c subject to regulation, and where

$$T^{cih} = \begin{cases} \pi_{cih} & \text{if } \theta_{ci} \leq \theta'_{ci} \\ 0 & \text{if } \theta_{ci} > \theta'_{ci}. \end{cases}$$

[18] Carve-outs are often made for policies that are "designed and applied" to protect legitimate public welfare objectives or for interventions that are not "manifestly excessive." The implementation of these concepts will depend on the context of the case in question. For instance, Annex 9-B.3(a) of the *Trans Pacific Partnership* agreement states: "The determination of whether an action or series of actions by a Party, in a specific situation, constitutes an indirect expropriation, requires a case-by-case, fact-based inquiry...". In our framework, such flexibility amounts to allowing different levels of investment protection for different industries and countries.

The first is whether countries serve as home and/or host countries, and the second is whether countries can make credible unilateral undertakings to protect foreign investment. To highlight the role of these basic premises for investment agreements in a compact way, we will focus on two archetypical settings.

The first type of agreement is between two fairly symmetric developed countries, such as TTIP, or the currently negotiated EU-Japan agreement. Such a "North-North" agreement is meant to stimulate investment flows in both directions. But the countries have well-developed domestic legal systems through which each party can unilaterally commit to any level of investment protection it finds desirable for the incoming foreign investment. To simplify the exposition, we will assume that the Northern countries are mirror images, implying that a North-North agreement specifies the same level of investment protection θ' for each country.

The other archetypical agreement is the traditional BIT between a developed and a developing country. Even if formally symmetric, the agreement serves to increase investments from the developed to the developing country only. The Southern partner cannot make credible unilateral commitments to protect incoming FDI. The investment protection that such a "North-South" agreement provides, effectively applies in the Southern partner only.

These two stylized agreements are likely to differ in an important regard: the symmetric interests of the parties to North-North agreements makes it reasonable to expect that they effectively *maximize the joint surplus of the contracting parties.* In contrast, as will be discussed further below, the parties to a North-South agreement have at least partly diverging interests with regard to investment protection, and such an agreement is therefore unlikely to maximize the joint surplus of the contracting parties.[19]

4 Properties of negotiated investment agreements

We assume that the terms of an investment agreement are determined through negotiations at the outset of the interaction. After the agreement is committed to, the same sequences occurs as above: investments are made, then the regulatory shock is realized, and finally the host country government decides whether to regulate. The agreement will impact the subsequent interactions through two broad channels. First, for a given investment level there will be a direct impact on the regulatory decision, since regulation will in certain cases now require compensation payments. Second, the commitment to compensation payments for certain situations will affect investment decisions, and this will in turn affect subsequent regulatory decisions.

Let θ' be the level of investment protection in an agreement based on the compensation function (10), and assume that this level is at least as high as the Nash equilibrium θ^0 to focus on the

[19] An additional reason why the negotiations over North-North, but not North-South, agreements could be expected to maximize joint welfare is that these involve not only investment protection, but also undertakings regarding trade liberalization, intellectual property right protection, and so forth. This creates possibilities for side payments that could overcome certain differences between the interests of the two countries.

non-trivial outcomes. Consider first the regulatory decision for a level of investment k. Assume that the host country must pay compensation if regulating. It then chooses between the welfare $V(\theta, k)$ from allowing production, and the welfare $-\Pi(k)$ from regulating. It will thus be indifferent for θ such that $V(\theta, k) = -\Pi(k)$. But this marginal value is of course $\Theta^J(k)$. Hence, due to the fact that the compensation function (10) requires *full* compensation whenever regulation is compensable, it aligns the incentives facing the host country with joint welfare incentives, for compensable shocks. This standard feature of IIAs will have very important implications for the efficiency properties of the agreement, as we will see.

For any given k, there will thus be either of two situations. If $\theta' < \Theta^J(k)$, the host country will abstain from regulating for $\theta < \theta' < \Theta^J(k)$, since regulation is then compensable and it is only willing to regulate with compensation for $\theta > \Theta^J(k)$. The host country will regulate for $\theta > \theta'$ since it will then not have to pay compensation. The expected welfare is then for given k

$$\int_{\underline{\theta}}^{\theta'} V(k, \theta) dF(\theta), \ \theta' \leq \Theta^J(k)$$

The other possibility is that $\Theta^J(k) < \theta'$. The host country will then abstain from regulating for $\theta < \Theta^J(k)$, it will regulate for $\Theta^J(k) < \theta < \theta'$ despite having to pay compensation, and it regulate without paying compensation for $\theta' < \theta$. The expected surplus of the host country for given k then equals

$$\int_{\underline{\theta}}^{\Theta^J(k)} V(k, \theta) dF(\theta) - [F(\theta') - F(\Theta^J(k)]\Pi(k), \ \theta' > \Theta^J(k),$$

where the second term reflects compensation payments.

Let us now turn to the investment decision. For any choice of k such that $\theta' < \Theta^J(k)$, the investor rationally expects no regulation for $\theta \leq \theta'$, and regulation without compensation for $\theta > \theta'$. The expected profit is thus $F(\theta')\Pi(k) - R(k)$. For any choice of k such that $\Theta^J(k) < \theta'$, the investor rationally expects no regulation for $\theta \leq \Theta^J(k)$, regulation with compensation $\Pi(k)$ for $\Theta^J(k) < \theta < \theta'$, and regulation without compensation for $\theta' < \theta$. The expected profit is thus again $F(\theta')\Pi(k) - R(k)$. The problem facing the investor is hence to maximize

$$F(\theta')\Pi(k) - R(k).$$

Let θ^E be the maximal level of investment protection for which the host country never pays compensation when regulating. This level is uniquely defined by $\theta^E - \Theta^J(K(\theta^E)) \equiv 0$ by the additional stability condition

$$\theta' < \Theta^J(K(\theta')) \text{ iff } \theta' < \theta^E. \tag{11}$$

By this assumption, an agreement with the level of investment protection limited to $\theta' \in [\theta^0, \theta^E)$ yields overregulation in the sense $\theta' < \Theta^J(K(\theta'))$, but the host country never has to pay any compensation. A more protective agreement—with $\theta' \in (\theta^E, \bar{\theta}]$—yields efficient regulation at $\Theta^J(K(\theta'))$, but compensation payments for all shocks in $(\Theta^J(K(\theta)), \theta')$.

4.1 The efficiency of an agreement

A fundamental question is whether the very simple investment agreement we have laid out above can resolve the overregulation/underinvestment problems in our model. One might expect the answer to be no because there is only one instrument, the level of investment protection θ', with which to offset the two distortions to investment and regulation. But as it turns out, the policy (10) can implement the jointly optimal outcome (θ^J, k^J) under a robust set of circumstances.

A North-North agreement In a North-North setting, the two symmetric countries negotiate the common level θ' to maximize the expected joint surplus:

$$\tilde{W}(\theta') \equiv \tilde{V}(\theta') + \tilde{\Pi}(\theta') = \int_{\underline{\theta}}^{\min\{\theta';\Theta^J(K(\theta'))\}} [V(K(\theta'), \theta) + \Pi(K(\theta'))]dF(\theta) - R(K(\theta')).$$

For $\theta' < \theta^E$, a marginal increase in investment protection has two first-order effects on the expected joint surplus:

$$\tilde{W}_\theta(\theta') = [V(K(\theta'), \theta') + \Pi(K(\theta'))]f(\theta') + X(K(\theta'), \theta')K_\theta(\theta').$$

There is a direct beneficial reduction of overregulation, captured by the first term. But more investment protection also triggers more investment, which can be beneficial or harmful, depending on the sign of the expected marginal externalities X. If $X(\theta^E, k^E) \geq 0$, increasing the level of investment protection θ' up to θ^E reduces the distortions to *both* regulation *and* investment.

For $\theta' \geq \theta^E$, a marginal increase in investment protection only affects the incentive to invest. The reason is that the full compensation requirement under the agreement causes the host country to internalize the full economic effects of its decision to regulate. Hence, the regulatory threshold will be the jointly optimal level $\Theta^J(K(\theta'))$. If full investment protection leads to overinvestment by the firm, then there exists a $\theta' \in [\theta^E, \bar{\theta}]$ such that $K(\theta') = k^J$.

We can thus characterize the equilibrium as follows (see Appendix A.2 for a proof):

Proposition 1 *The equilibrium level of investment protection θ^N and associated investment $k^N = K(\theta^N)$ in a North-North agreement based on the compensation function (10):*
(i) implements the unconstrained jointly optimal outcome (θ^J, k^J) if $X(k^E, \theta^E) \geq 0$ and full investment protection would lead to overinvestment ($k^J \leq K(\bar{\theta})$);
(ii) yields overregulation, $\theta^N < \Theta^J(k^N)$, and overinvestment, $k^N > K^J(\theta^N)$, if $X(k^E, \theta^E) < 0$.

Hence, a North-North agreement can implement the unconstrained joint optimum under a robust set of circumstances, despite the simplicity of the compensation function (10). When this is infeasible, the equilibrium agreement entails a trade-off between overregulation and overinvestment.

We claimed above that the focus on agreements with investment-independent investment protection levels θ' does not restrict the scope of an agreement. This is clearly the case with a non-negative investment externality ($X(k^E, \theta^E) \geq 0$), since the first-best is then implemented. But could an agreement with an investment protection level $\hat{\Theta}(k)$ that depends on k, achieve something better if the externality is negative? To see why not, note that as long as investors take the probability of regulation as unaffected by their own investment decisions, their investments will only depend on the level of investment protection, be it θ' or $\hat{\Theta}(k)$. Hence, by setting $\theta' = \hat{\Theta}(k)$, the simpler agreement can implement whatever investment that results from the more complex agreement. Furthermore, the regulatory decision also only depends on the level of investment protection, and on the level of investment. Consequently, the outcome could not be improved upon by using a more general investment protection level $\hat{\Theta}(k)$ in the present setting. But Section 4.9 will point to a setting where the more general form of agreement would be highly useful.

A North-South agreement North prefers full protection ($\theta' = \bar{\theta}$) of its investment in South because the expected profit of FDI is strictly increasing in the level of investment protection in South. South's fall-back position is the Nash equilibrium (θ^0, k^0) as it cannot make credible unilateral commitments to protect investment absent an agreement. Hence, there will be scope for a North-South agreement if the marginal externalities are sufficiently large: $\tilde{V}_\theta(\theta^0) = X(k^0, \theta^0) > 0$. The equilibrium level of investment protection θ^S will be somewhere in the interval $(\theta^0, \bar{\theta}]$ depending on the two countries' relative bargaining strengths. If North can dictate the terms of the agreement, the equilibrium will be the maximal level acceptable to South, i.e. the largest θ^S and associated investment $k^S = K(\theta^S)$ satisfying $\tilde{V}(\theta^S) \geq \tilde{V}(\theta^0)$.

We saw that a North-North agreement under certain circumstances can handle both the regulatory and investment distortions. But in the case of a North-South agreement, there is an additional source of distortion: The contracting parties' have conflicting interests about the optimal level of investment protection because of how it distributes the surplus between them. Hence, there is no presumption that the outcome would be jointly efficient, regardless how the contract terms are negotiated. In fact, a jointly desirable agreement could fail to materialize at all. This will be the case if the marginal externalities are negative, $X(k^0, \theta^0) < 0$, but the value to foreign investors of investment protection is sufficiently large that

$$\tilde{W}_\theta(\theta^0) = X(k^0, \theta^0) K_\theta(\theta^0) + f(\theta^0) \Pi(K(\theta^0)) > 0.$$

Consequently:

Proposition 2 *A North-South agreement will be formed if $X(k^0, \theta^0) > 0$. When formed, there is no presumption that the agreement will implement a joint welfare maximizing level of investment protection. If not formed, an agreement could still increase joint welfare.*

Our findings here suggest one possible reason why developing countries renegotiate their IIAs, or redraft their model BITs: as these countries become more developed, and more important as source countries for FDI, they become more equal bargaining partners to developed countries. They will therefore have the incentives and ability to revise their agreements in order to increase the carve-outs for their domestic regulations.

Two rationales for investment agreements The stylized North-South and North-North scenarios identify two separate roles that investment agreements might play. The welfare benefits from a North-South investment agreement here stem entirely from the credibility the agreement lends to South's commitment to compensate for regulation for $\theta > \theta^0$. If South had full unilateral commitment possibilities, it would choose $\theta^U \equiv \arg\max_{\theta'} \tilde{V}(\theta')$ absent any agreement, in which case there would be no gains from an agreement: $\tilde{V}(\theta') - \tilde{V}(\theta^U) \leq 0$ for all θ'. The role of North-South agreements thus corresponds closely to the notion that trade agreements serve as commitment devices, helping governments to withstand domestic protectionist pressures.[20] But 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, since it will have to share the surplus with North.

The benefits of a North-North agreement instead arise from *internalization of negative international externalities* from national regulatory policies. With investments flowing in both directions, the parties can negotiate improved investment protection abroad by offering improved investment protection for foreign investment at home. This corresponds closely to the standard view of the gains from trade agreements, which sees these agreements as means for taking countries out of Prisoners' Dilemmas by allowing them to exchange increased imports for increased exports to mutual benefit.

Observation 3 *The rationale for a North-South agreement is South's lack of unilateral commitment possibilities regarding investment protection. 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 can substitute for commercial contracts between host countries and individual investors. In commercial contracts the parties can contract the level of investment, and possibly also regulatory policies. Such contracts would be superior to state-to-state agreements for Southern countries (disregarding transaction costs). Commercial contracts cannot always replace North-North agreements however, since the negotiations over commercial contracts typically do not allow for exchanges of concessions.

Observation 4 *Contracting directly on investment and regulation with individual investors would*

[20] 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, but other forms of externality problems.

be better than a North-South agreement (absent transaction costs), but would not necessarily dominate a North-North agreement.

4.2 Two desirable features of the compensation scheme

The compensation scheme (10) reflects basic legal principles concerning state responsibility and sovereignty. But it might appear exceedingly rigid from an economic perspective by offering full compensation for all regulatory shocks below a threshold value, but no compensation for more severe shocks. Is there here a conflict between legal and economic principles?

The threshold property The economic desirability of using a threshold for compensable regulation is not (at least us) as self-evident as it might seem. For instance, could it not be optimal to request compensation for large θ as a distortion-free way of stimulating investment, as the host country will anyway regulate for these values? Also, the agreement is designed prior to the realization of the shock, so every possible outcome must be weighted by the density $f(\theta)$ to obtain the expected net benefit of allowing production. But we have hardly imposed any structure on the distribution of the shocks θ. Hence, the ex ante optimal compensation scheme could in principle yield non-monotonic regulation in θ, depending on the properties of $f(\theta)$. So how can we exclude that it could be optimal to request compensation for certain ranges of θ but not for other, if we have complete freedom to design any complex compensation function? We prove the following result in Appendix A.3:

Proposition 3 *For any agreement satisfying the salient features (a)-(c) there exists an alternative agreement satisfying the same restrictions, and that for each host country:*
(i) implements a threshold function for regulation $\hat{\Theta}(k) \in [\Theta(k), \Theta^J(k)]$*; and*
(ii) yields weakly higher expected domestic welfare and foreign industry profits than under the initial agreement.

The proof of the Proposition is constructive, showing how one can replace any initial compensation scheme satisfying features (a)-(c) with another scheme that has the properties stated in the Proposition. This alternative compensation scheme is a convex combination of the firm's foregone operating profit and the payment under the initial compensation scheme.[21] Hence, the alternative scheme relies only on the features of the initial scheme and on foregone operating profits.[22]

Proposition 3 thus establishes the optimality of a threshold for when there will be regulation. The Proposition applies to a much broader set of circumstances than the single investor setting we employ here. It holds for Pareto efficient agreements more generally, not only those that maximize

[21] 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 host country i.

[22] For instance, the alternative scheme does not involve draconian punishments for deviations from some desirable outcome, if this is not part of the initial scheme.

joint welfare. It holds for compensation schemes that pay less than foregone profits. It holds in the presence of non-discrimination clauses, with asymmetric firms, imperfect competition, and mixed foreign/domestic ownership structures, and the Proposition also holds when firms invest strategically to influence regulatory decisions.

A useful implication of Proposition 3 is that we often do not need to consider complex investment agreements, we can focus on agreements with a threshold value for compensation.

The full compensation property A second rigid feature of the compensation scheme is that it requests either full or zero compensation for foregone operating profit. Can this simplistic feature explain the inefficiency of North-North agreements for negative expected marginal externalities? The following Proposition, which we prove in Appendix A.4, shows that it would require a qualitatively very different type of compensation scheme to improve on the outcome:[23]

Proposition 4 *The compensation function (10) maximizes the joint expected surplus of the contracting parties in the class of compensation functions that are non-decreasing in investment.*

The requirement of full compensation for foregone operating profits not only reflects a basic principle in international law regarding state responsibility. It also has the economically desirable property of effectively inducing host countries to internalize all ramifications of their regulatory decisions, thus aligning host country incentives with the ex post efficient level of regulation. Proposition 4 further establishes that foregone operating profit should form the basis of all compensation under the restriction that firms cannot be compensated less if they have more to lose from being regulated. The proposition also shows that compensation would have to be a decreasing function of profit in order to induce efficient investment if the marginal investment externalities are negative.

4.3 Compensation payments need not indicate bad-faith regulation or litigation

Instances where investors successfully litigate against host countries and receive compensation are often seen to indicate that host countries have deliberately violated the letter or spirit of IIAs, or alternatively as indications of flaws in the legal regime that allow investors to extract protection they should not receive. These are not the only possible explanations, as our framework suggests.

Under a North-North agreement with positive expected marginal investment externalities, the equilibrium level of investment protection will be high enough, $\theta^N > \theta^E$, that it is ex optimal for the host country to regulate for $\theta \in (\Theta^J(k^N), \theta^N]$ despite having to pay compensation. In the case of a North-South agreement, North chooses the maximal level of investment protection South is willing to accept, if North has most of the bargaining power. This level exceeds θ^E if $\tilde{V}(\theta^E) > \tilde{V}(\theta^0)$. But in could also more generally be in host countries' interest to accept compensation payments.

[23] The proof of Proposition 3 establishes the optimality of full compensation for θ below the established threshold, but does not show the optimality of zero compensation for higher θ.

This happens if the marginal investment externalities are sufficiently positive that $\theta^U > \theta^E$. In this case, the equilibrium investment agreement features compensation payments in the host country independently of trade flows and the distribution of bargaining power in the negotiation.

Corollary 1 *Compensation payments occur in equilibrium for shocks in the range* $(\Theta^J(K(\theta')), \theta']$ *if the expected marginal production externality is sufficiently positive, and/or* $\theta' > \theta^E$.

Hence, compensation payments—and probably also litigation, since investors in practice often have to litigate in order to receive compensation payments—*do not necessarily indicate flaws in the IIA regime*.[24] Indeed, compensation payments are essential in achieving the unconstrained joint optimum (k^J, θ^J) in a North-North agreement when there are positive expected marginal externalities. Note that the role of the payments then is not to prevent overregulation, but to stimulate investment, the benefit of which materializes for *other* realizations of θ than those for which the compensation payments are made. Compensation payments then serve as implicit subsidies to investment.

4.4 Regulatory chill

A core claim in the policy debate holds that investment agreements cause *regulatory chill*. The notion is rarely precisely defined, but can be given several different interpretations within the context of our model.[25] We will take *domestic* regulatory chill to mean that an agreement prevents host countries from undertaking regulations that are ex post unilaterally desirable absent compensation requirements. A corresponding *joint* regulatory chill occurs if the agreement induces host countries to allow production in situations where regulation would have been jointly optimal ex post. The following result is a direct implication of Proposition 3:

Corollary 2 *A Pareto-efficient agreement implements ex post efficient production for* $\theta \leq \Theta(k)$, *ex post inefficient regulation for* $\Theta(k) < \theta < \Theta^J(k)$, *and ex post efficient regulation for* $\theta \geq \Theta^J(k)$). *There will hence be domestic, but not joint, regulatory chill.*

To see the generality of the underlying argument regarding joint regulatory chill, note first that joint regulatory chill can occur only if the agreement stipulates compensation in excess of foregone operating profits for some $\theta > \Theta^J(k)$. The profit is $\Pi(k)$ for such realizations of θ, since production is allowed in case of regulatory chill. Reducing compensation for regulation to $\Pi(k)$ for these values of θ would induce the host country to regulate, but would not affect investment incentives as the firm would still receive its operating profit for those shocks-albeit now as compensation for regulation. The modification of the compensation scheme thus increases regulatory efficiency by eliminating joint regulatory chill without influencing investments or profits, and therefore represents a Pareto improvement.

[24] A more complete argument would take account of litigation costs.

[25] Janeba (2016) is an exception.

It is easy to see how the domestic regulatory chill that will arise for $\theta \in (\Theta(k), \Theta^J(k))$ might appear as a failure of the agreement: Production is undesirable from a national perspective for this range of θ, and it is the threat of having to compensate that de facto deters the host country from regulating. For these shocks the agreement actually redistributes surplus from domestic interests to foreign investors. But this reasoning disregards two important aspects. First, production is still *jointly* desirable for these θ since there is no joint regulatory chill. Second, the protection that the agreement provides for these regulatory shocks *stimulates investment* that is desirable to the host country for weaker shocks.

The argument above is robust in the sense of being valid for a broad range of market and ownership structures. It also holds for *any* k. But its welfare implications rest on the presumption that the objective function of the host country corresponds with domestic welfare and therefore is the relevant criterion for evaluating the efficiency of the agreement. Assume instead that host country welfare is given by $V(k, \theta)$ whereas the host country government obtains utility $V(k, \theta) + \alpha\Pi(k)$ if allowing production. We let $\alpha\Pi(k) > 0$ represent the share of operating profits the government can privately extract, for instance in the form of bribes or campaign contributions. The net operating profit of the firm then is $(1 - \alpha)\Pi(k)$.[26] Importantly, these payments are not a source of welfare from an overall perspective. Furthermore, the host country is not able to extract any rent from compensation payments under the investment agreement. With these modifications, $\Theta^J(k)$ still characterizes the ex post efficient level of regulation. The government threshold for regulation $\Theta^G(k, \alpha)$ is defined by

$$V(k, \Theta^G) + (1 + \alpha)\Pi(k) \equiv 0$$

if the host country must pay the full foregone operating profit as compensation for regulation. The rent extraction motive yields an incentive to underregulate from an efficiency viewpoint: $\Theta^G(k, \alpha) > \Theta^J(k)$ if $\alpha > 0$, which we for simplicity take as exogenously given. The threshold $\Theta(k, \alpha)$ for regulation is defined by

$$V(k, \Theta) + \alpha\Pi(k) \equiv 0$$

if the host country does not have to pay any compensation.

We also need to consider investment incentives to determine whether there is joint regulatory chill in equilibrium. The expected operating profit is $F(\hat{\theta})(1-\alpha)\Pi(k) + [F(\theta') - F(\hat{\theta})]\Pi(k)$, where θ' is the level of investment protection. The expected profit depends also on the anticipated threshold for regulation $\hat{\theta}$, since we assume that the host country is only able to extract rent from production, but not from compensation. The equilibrium level of investment $K(\theta', \hat{\theta}, \alpha)$ is thus given by the

[26] An example of the importance of anti-corruption in the context of IIAs is Art. 17 of the recently negotiated (but still not ratified) BIT between Morocco and Nigeria (available at investmentpolicyhub.unctad.org/Download/TreatFile/5409). The provision explicitly forbids investors to "...offer or give any undue pecuniary or other advantage..." with the purpose of extracting beneficial treatment, and it obliges the parties to enforce their laws in this regard.

solution to

$$[F(\theta') - F(\hat{\theta})\alpha]\Pi_k(K) - R_k(K) = 0.$$

Let $\theta^0(\alpha)$ be defined by $\theta^0 - \Theta(K(\theta^0, \theta^0, \alpha), \alpha) \equiv 0$, $\theta^E(\alpha)$ by $\theta^E - \Theta^J(K(\theta^E, \theta^E, \alpha)) \equiv 0$, and assume that the stability conditions (6) and (11) extend to $\alpha > 0$. The divergence between government and host country national interests implies that ex post inefficient underregulation can occur under certain circumstances (the proof is in Appendix A.5):

Proposition 5 *Assume that the host country government appropriates a share $\alpha > 0$ of operating profit by allowing production, and that the compensation scheme is given by (10).*

(i) There will be no joint regulatory chill if $\theta' \in [\theta^0(\alpha), \theta^E(\alpha)]$.
(ii) There will be joint regulatory chill for $\theta \in (\Theta^J(K(\theta', \hat{\theta}, \alpha)), \min\{\theta'; \Theta^G(K(\theta', \hat{\theta}, \alpha), \alpha)\}]$ *if* $\theta' > \theta^E(\alpha)$.

This is obviously a simplistic depiction of for instance corruption. The aim is just to point to a mechanism that could lead to joint regulatory chill.

4.5 Winners and losers

A fundamental issue for the politics of investment agreements is the distribution of the costs and benefits across the population. The source country unilaterally prefers full protection, $\theta' = \bar{\theta}$, for its outgoing investment, whereas the host country unilaterally prefers the level of investment protection that maximizes its expected domestic welfare, $\theta' = \theta^U$.

South is unable to make credible unilateral commitments to any other level of investment protection than the Nash Equilibrium θ^0. As we saw above, there will be scope for a North-South agreement if the expected marginal investment externalities are positive absent any agreement: $X(k^0, \theta^0) > 0$. South would then ideally increase the level of investment protection to θ^U, but not further. And because Northern expected welfare increases monotonically in the level of investment protection in South, this is the minimum level of protection that a North-South agreement will entail. If both parties wield some bargaining power, the equilibrium θ^S will be strictly above θ^U, but still low enough that domestic welfare in South is higher than absent an agreement: $\tilde{V}(\theta^S) > \tilde{V}(\theta^0)$. However, the net gain $\tilde{V}(\theta^S) - \tilde{V}(\theta^0)$ will be small if North has most of the bargaining power.

In contrast, due to the symmetry between the parties, all gains from a North-North agreement will be internalized. But these gains come with pronounced distributional implications. As each country can unilaterally implement its preferred protection level for incoming FDI, it will suffer a loss $\tilde{V}(\theta^U) - \tilde{V}(\theta^N) > 0$ in domestic welfare because of the agreement on $\theta^N > \theta^U$.[27] The sole

[27] If $\theta^U \in (\theta^0, \bar{\theta})$, then the host country and the source country both benefit from increasing θ' to θ^U for any $\theta' < \theta^U$. A marginal increase in θ' above θ^U has a first-order effect on industry profit, $\tilde{\Pi}_\theta(\theta^U) > 0$, but only a second-order effect on domestic expected welfare, $\tilde{V}_\theta(\theta^U) = 0$. Setting $\theta^N > \theta^U$ thus increases total surplus. If $\theta^U = \theta^0$, then an agreement is economically meaningful only if $\theta^N > \theta^0 = \theta^U$; see Appendix A.6. If $\theta^U = \bar{\theta}$ there are no gains from an agreement since the host country already offers maximal protection.

reason for accepting such a loss is that investment protection in the partner country increases to such an extent that the expected profit increase on outgoing FDI dominates the loss in domestic welfare: $\tilde{\Pi}(\theta^N) - \tilde{\Pi}(\theta^U) > \tilde{V}(\theta^U) - \tilde{V}(\theta^N)$.

Proposition 6 *(i) The negotiated level of investment protection, and the level of investment, will exceed the level that maximizes the respective host country domestic welfare.*
(ii) A North-South agreement benefits Northern investors and increases domestic welfare in South.
(iii) A North-North agreement benefits foreign investors in both countries, but reduces expected domestic welfare in both countries.

Note that (i) does not imply that host country domestic welfare will fall as a result of an agreement, only that the level of investment protection will be too high from its domestic perspective. This is domestic regulatory chill. The investment agreement will thus *always appear too protective of foreign investor interests, from the point of view of rest of society.*

We believe that these observations shed light on the policy debate regarding investment agreements. The costs and benefits for the Southern parties to North-South investment agreements have been discussed for years. But generally speaking, several thousands of such agreements were signed without much political upheaval. This contrasts sharply with the heated debate concerning the attempts to include investment protection in more symmetric agreements, and most notably in CETA, TPP, and TTIP. Existing legal systems in e.g. the EU and the U.S. are likely to provide sufficient protection of FDI to internalize domestic welfare effects. The *additional* investment protection that these agreements would offer would thus mainly benefit foreign investors, but harm the rest of society, for instance by exacerbating domestic regulatory chill. Incidentally, these distributional effects appear closely compatible with arguments that have been put forward by the U.S. Administration and the EU Commission as to the benefits of investment protection in TTIP. Both sides have emphasized the benefits from increased protection of their respective *outgoing* investment flows, but have rarely pointed to benefits from increased domestic investment protection.

4.6 Non-discrimination

Critics of IIAs sometimes claim that their *sole* role should be to prevent discriminatory treatment of foreign investment. For instance, Stiglitz (2008, p. 249) argues that "...non-discrimination provisions will provide much of the security that investors need without compromising the ability of democratic governments to conduct their business." We will here examine how far a National Treatment provision would take the parties toward solving the investment and regulatory distortions.

One obvious limitation is that a National Treatment clause will be ineffective absent national firms that operate under sufficiently "like circumstances" to those facing foreign investors. We therefore focus on the case where each host country features 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, suffer from the same country-specific shock, and therefore produce under "like circumstances" for the purpose of an NT provision. But the sectors are economically unrelated to avoid that strategic considerations influence regulatory 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.

Absent an agreement, and absent the ability to make unilateral commitments to investment protection, the host country will regulate the foreign industry more frequently than the domestic industry: $\theta_F^0 = \theta^0$ and $\theta_D^0 = \theta^E > \theta^0$ (see Observation 1). The host country will choose different levels of protection in the two sectors also if it is able to make a unilateral commitment; we then have $\theta_F^U = \theta^U$ and $\theta_D^U = \theta^W > \theta^U$, where $\theta^W \equiv \arg\max_{\theta' \geq \theta^0} \tilde{W}(\theta')$.[28]

A foreign firm here faces two forms of discrimination. First, the host country regulates foreign investment more frequently, despite both industries being subject to the same shock: $\theta_F^0 < \theta_D^0$ and $\theta_F^U \leq \theta_D^U$. Second, for any given k the host country regulates industry F more frequently, since $\Theta(k) < \Theta^J(k)$. This will not have any separate implications however, as long as firms treat regulatory decisions as unaffected by their respective investment choices. So what impact would a commitment to NT have?

We represent an NT clause by the requirement that $\theta_F \geq \theta_D$, where the weak inequality reflects the fact that NT clauses do not rule out a comparatively more favorable treatment of foreign investment. This restriction ensures that the domestic industry is regulated whenever there is regulation of the foreign industry, but not vice versa. To limit the number of cases to consider, we focus on the North-North and North-South settings.

An agreement comprising NT only With a North-North agreement on NT only, the unilaterally determined common protection level θ_{NTOnly}^N has the following features (see Appendix A.7 for a proof):

Lemma 2 $\theta^U \leq \theta_{NTOnly}^N \leq \theta^W$, *with strict inequalities if* $\theta_{NTOnly}^N \in (\theta^0, \bar{\theta})$.

Consequently, this agreement increases (reduces) the level of investment protection for foreign (domestic) investors beyond (below) the unilaterally optimal level and therefore reduces host country domestic welfare. With the two-way investment flows, both countries benefit from the better treatment of their foreign investment, but it is ambiguous whether they would accept such an agreement. Indeed, it might be better with *no agreement* compared to an agreement that only imposes NT.

Under a North-South agreement comprising NT only, South would unilaterally choose the common investment protection level θ_{NTOnly}^S for its foreign and domestic sectors, as given by

$$2V(K(\theta_{NTOnly}^S), \theta_{NTOnly}^S) + \Pi(K(\theta_{NTOnly}^S)) \equiv 0$$

[28] The associated investment levels are $k_F^0 = k^0$, $k_D^0 = k^E > k^0$, $k_F^U = K(\theta^U)$ and $k_D^U = K(\theta^W) > k^U$.

The first term is the domestic welfare derived from the two industries, and the second term the profits of the domestic industry. It is easy to verify that $\theta^S_{NTOnly} \in (\theta^0_F, \theta^0_D)$ under a similar stability condition as (6). Consequently, there will be increased foreign investment, and reduced domestic investment. This benefits North, but has ambiguous implications for domestic welfare in South.

Proposition 7 *The imposition of only an NT provision:*
(i) has ambiguous welfare consequences in a North-North agreement;
(ii) benefits North, and has ambiguous consequences for South, in a North-South agreement.

NT as a complement to protection against regulatory expropriation How would the presence of NT affect the optimal design of investment protection? A North-North agreement will yield the same efficient protection level in the two partners' respective F sectors. By symmetry across industries, this is exactly the same level of investment protection as in the respective domestic industry absent an NT clause: $\theta_D = \theta_F = \theta^W = \theta^N$. Consequently, there is no role for NT here.

In a North-South setting, NT again has an impact only when there is less protection of foreign than domestic investment absent NT: $\theta^S < \theta^E$. Let θ^S_{NT} be the negotiated level of investment protection in a Pareto optimal North-South agreement under NT. It will be optimal for South to set $\theta^S_{NT} \in (\theta^S, \theta^E)$. This will be better for North than the level θ^S, since it offers more protection. But it will have ambiguous consequences for South. The host country welfare in sector D (including the profits of the domestic industry) can potentially increase if the equilibrium level of investment protection moves closer to the unilaterally optimal level θ^W than before. However, the host country welfare from the industry with incoming foreign investment is likely to fall, because the NT clause here increases the level of investment protection even further away from the domestically optimal level θ^U.

Proposition 8 *The imposition of an NT provision as a complement to undertakings on investment protection:*
(i) serves no purpose in a North-North agreement; and
(ii) either serves no purpose, or benefits North and has ambiguous welfare consequences for South, in a North-South agreement.

Note that the role of NT here is very different from that normally played by NT in trade agreements, where NT is seen as an instrument to prevent parties to an agreement from opportunistically exploiting contractual incompleteness and thereby undermining bargaining concessions. NT renders such opportunistic behavior less attractive by effectively forcing the importing country to distort also its domestic production if it wants to distort trade.[29] The purpose of NT is different in this setting, as it is not meant to neutralize opportunistic behavior—there is no commitment that is

[29] See Horn (2006), Saggi and Sara (2008), and Horn et al (2010).

eroded due to the complete economic separation between the two sectors. Instead, NT here essentially serves to extend the commitment possibilities that the investment agreement brings to the domestic sector

Observation 5 *NT allows countries that lack credible unilateral commitment possibilities to indirectly use the enforcement mechanism offered by investment agreements to solve ex post underregulation problems in their domestic sectors.*

Note finally that NT could play a more substantial role if there were other policy instruments host countries could use to undermine undertakings concerning investment protection levels.

4.7 Carve-outs for direct expropriations

Direct expropriations were particularly common during the 1950s and 1960s, but are not entirely something of the past (Hajzler, 2012). IIAs typically have stricter rules regarding compensation for direct than indirect expropriation in the sense that carve-outs from compensation requirements only apply to regulatory expropriations. This might have intuitive appeal as direct expropriations can be seen as opportunistic take-overs of rents belonging to foreign investors. But direct expropriations typically are defended as legitimate exercises of national sovereignty, and claims for compensation are often contested. Surprising perhaps, the latter argument can be given some economic support.

Consider an investment agreement of the form analyzed above, with a level of investment protection $\theta' < \theta^E$ and equilibrium investment $k' = K(\theta')$. Under this agreement, there is inefficient overregulation for $\theta \in (\theta', \Theta^J(k'))$. Compare this with an alternative agreement that is identical in all respects, except for allowing direct expropriation without compensation for $\theta \in (\theta', \Theta^J(k')]$. Investors would be indifferent, since they would anyway only receive payments for $\theta \leq \theta'$, and would therefore invest k' under both agreements. But under the alternative agreement, the host country now allows production for $\theta \in (\theta', \Theta^J(k')]$, and it is therefore more efficient. These beneficial features of direct expropriation can actually take us very far (the proof is found in Appendix A.8):

Proposition 9 *The unconstrained joint optimum (θ^J, k^J) can be implemented as a Nash equilibrium through an international investment agreement that allows direct expropriation if expropriation does not reduce the profits from the expropriated assets.*

The point is not to argue that uncompensated direct expropriation should be allowed in actual agreements, but rather that the reason for disallowing direct expropriation is less trivial than it might seem. It is precisely the fact that these measures constitute a pure transfer of ownership that provides a role for them in investment agreements. Their purpose is then to mitigate overinvestment to arise from the full compensation requirement that in turn serve to mitigate the ex post incentive to regulate.

An important caveat for the above argument is that θ has to be observable and verifiable. Assume instead that the host country has private information about θ. The value of allowing production in the foreign-owned industry is $V(k,\theta)$, whereas the value of direct expropriation is $V(k,\theta)+\Pi(k)-T^x$, 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 investment agreement can ensure that foreign ownership is 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 were reversed, 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 in our setting it impossible to have direct expropriation for some realizations of the shock and regulation for other realizations if the host country is privately informed about the shock. Instead, the host country has to choose either private or 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: $T^x = \Pi(k)$.

4.8 Insurance against political risks

The shocks that we have considered above have been interpreted to represent the arrival of information that affects the desirability to a particular national decision maker of regulating production. But regulatory interventions could also be triggered by changes to the preferences of decision makers, for instance, because of changes of government. It is a contested issue whether such "political" shocks should be compensable. Some argue that investors should expect political changes, and that such shocks therefore should not be compensable.

To shed a little light on this thorny issue, we make a simple reinterpretation of the model, by assuming that the factual magnitude of the regulatory problem is known, but that governments differ in their sensitivity θ to the damage, with higher θ representing higher sensitivity. Let $V(k, \theta)$ be the ex post welfare of a government with preference parameter θ. We assume that the agreement is negotiated at the outset behind a veil of ignorance regarding the regulatory preferences of the government that will be in place after the investments have been made, with expectations taken over θ. Investments are made after the agreement is in place, but before the identity of the government is determined. This setting is formally identical to the one above, the only difference being the interpretation of the model:

Observation 6 *If an agreement is negotiated behind a veil of ignorance regarding host country regulatory preferences, regulation without compensation should be allowed if and only if the government is sufficiently sensitive to the regulatory problem.*

This reasoning obviously disregards the incentives for incumbent governments to design agreements so as to influence future regulatory decisions. But our point is not to argue that political risks should be treated any particular way, only that is not self-evident from an economic perspective that political risks should not be compensable.

4.9 Strategic investors

We have thus far assumed that investors do not seek to affect the probability of regulation, as seems plausible when their respective contribution to the regulatory problem is small. But a large investor might plausibly take account of how the investment affects the probability of regulation. This raises the question of whether an investment agreement will perform worse with such "regulation-strategic" investors, or perhaps even be harmful to the host country. As we will see, the effect will be just the opposite.

Assume that there is a single, policy-strategic investor in an industry. Absent an investment agreement, it will invest to maximize $F(\Theta(k))\Pi(k) - R(k)$, where the difference compared to the non-strategic investor hence is that the investor here takes account of the impact of its investment on $F(\Theta(k))$. Let

$$\pi^{SPE} \equiv \max_k F(\Theta(k))\Pi(k) - R(k)$$

be the resulting sub-game perfect equilibrium profit, $k^{SPE} > 0$ the firm's optimal investment and $\theta^{SPE} = \Theta(k^{SPE})$ the equilibrium level of regulation. Assume that these are unique.

Consider next the outcome with an agreement. We have assumed for expositional reasons that the threshold for compensation is independent of the level of investment. But with a regulation-strategic investor, such a restriction would affect the outcome. We therefore allow for the investment protection level to depend on the level of investment. The following finding is established in Appendix A.9:

Proposition 10 *Assume that the industry consists of a single regulation-strategic foreign firm. Then there exists an investment protection function $\tilde{\Theta}(k)$ and compensation function*

$$\tilde{T}(\theta, k) \equiv \begin{cases} \Pi(k) & \text{for } \theta \leq \tilde{\Theta}(k) \\ 0 & \text{for } \theta > \tilde{\Theta}(k) \end{cases} \tag{12}$$

that implement the unconstrained jointly optimal outcome (θ^J, k^J) as a sub-game perfect equilibrium, if and only if $\Pi(k^J) - R(k^J) \geq \pi^{SPE}$.

Hence, an agreement can now implement the jointly optimal outcome even if the marginal investment externality is *negative*. To see why, assume that $X(k^E, \theta^E) < 0$, so that the firm tends to overinvest. Since the firm's investment decision is affected not only by the level of investment protection, but also by the marginal change in investment protection as a function of the investment, the tendency toward overinvestment can be mitigated by letting the level of investment protection decrease in the level of investment: $\tilde{\Theta}'(k) < 0$.

The profit condition in Proposition 10 is a participation constraint. If it does not hold, it is better for the source country to reject a proposed investment agreement rather than to accept an agreement that implements k^J, even if the agreement goes as far as to offer complete investment protection. The only way to implement the joint optimum is then to compensate the firm in excess of its operating profit. But if such punitive damages are not allowed, the contracting parties cannot under any circumstances benefit from agreeing on any other compensation mechanism than a carve-out with full compensation (the proof is in Appendix A.10):

Proposition 11 *Assume that the industry consists of a single regulation-strategic foreign firm. Assume also that investment agreements at most can award foregone operating profit in compensation. Any Pareto optimal outcome can then be implemented by means of an investment protection function* $\tilde{\Theta}(k)$, *and compensation function (12).*

Proposition 11 shows that a simple carve-out policy (12) can be optimal even if the conditions are such that the parties cannot implement the jointly efficient solution (θ^J, k^J). Such a situation could arise e.g. if the participation constraint in Proposition 10 is violated, or the negotiations do not maximize joint surplus. The qualifying condition is that agreements are constrained to compensate investors at most for their foregone operating profit.

5 Going beyond standard compensation schemes

The compensation scheme (10) can fully correct both the investment and regulatory distortions for a joint welfare maximizing agreement if the marginal investment externalities are non-negative. But when these externalities are negative, the agreement must deviate from some of features (a)-(e) of actual IIAs to implement an efficient outcome. In what follows, we first review a number of compensation schemes that have been identified in the literature, and that would do the job in our framework. We then present an alternative efficient scheme, building on relative performance measures. We finally consider how an agreement must be modified if the host country has private information concerning regulatory shocks and the industry is a monopoly.

5.1 Schemes in the literature

The seminal contributions to the Law and Economics literature on takings were made by Blume et al. (1984), and Miceli and Segerson (1994).[30] The early literature implicitly assumed undistorted incentives to invest and to regulate, in which case a compensation mechanism can only reduce welfare. But Hermalin (1995) considered distortions to investments and regulation in a model with

[30] See Miceli and Segerson (2011) for a comprehensive survey.

direct expropriation and a single firm. He derived two efficient mechanisms. In the first, a firm pays a production tax equal to the country's value of seizing the asset. A tax equal to $-V(k,\theta)$ would implement the fully efficient outcome in our setting, since the firm would then internalize the full social cost of its actions. The second mechanism requests the host country to pay the firm the same amount as compensation subsequent to expropriation. This feature highlights the fundamental property of efficient compensation that it should be based also on the social cost and not only on operating profits. Even so, Hermalin's (1995) second compensation rule is inefficient under the threat of indirect expropriation. It yields an expected compensation of $-\int_{\theta^J}^{\bar{\theta}} V(k,\theta)dF(\theta)$ in the current setting, which generally differs from the expected compensation $-\int_{\underline{\theta}}^{\theta^J} V(k,\theta)dF(\theta)$ that generates efficient investment incentives. Efficiency here requires the investor to internalize the marginal effect of the investment on the expected social value of *allowing* production rather than the value of *disallowing* it.

Following Blume et al. (1984), Aisbett et al. (2010a) study a deviation from feature (e) of IIAs, in which compensation is a linear combination of operating profits and investment costs: $T(k) = \delta\Pi(k) + \beta R(k)$. It is shown how full efficiency can be achieved also with distorted incentives to regulate if the host country can overcompensate the industry for regulation. This compensation mechanism has two instruments (δ, β) that can be used to correct the distortions to investment and regulation within our framework. In our setting, it is possible to implement the fully efficient outcome (k^J, θ^J) as a Nash equilibrium with a special case of this compensation scheme where

$$T(k,\theta) = \begin{cases} \Pi(k) & \text{if } \theta \leq \theta^J \\ \delta\Pi(k) + (1-\delta)\frac{\Pi(k^J)}{R(k^J)}R(k) & \text{if } \theta > \theta^J. \end{cases} \tag{13}$$

A virtue of (13) is that implementation does not require the shock θ to be publicly observable.[31] However, the compensation scheme deviates from those in actual IIAs in several respects. First, there are no strict carve-outs since compensation is paid for any regulation. Second, the scheme in (13) overcompensates firms for their losses since $T(k) > \Pi(k)$ for some investments $k \neq k^J$.[32] Also, it requires that it is feasible to estimate the firm's operating profit $\Pi(k^J)$ and capital cost $R(k^J)$ at the efficient investment level k^J, which would presumably make its practical implementation difficult. Furthermore, the compensation rule (13) requires that δ is firm-specific, 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 information regarding k^J, and that does not request sym-

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

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

metric firms. Adapted to our setting, the compensation

$$T(k) = \frac{\widetilde{T} + \int_{\underline{\theta}}^{\Theta^J(k)} V(k,\theta) dF(\theta)}{1 - F(\Theta^J(k))} \tag{14}$$

induces efficient investment if regulation is ex post efficient, i.e., if the host country applies the regulatory threshold $\Theta^J(k)$, even if firms behave strategically. In particular, $T(k)$ only depends on the actual investment k. Ex post efficient regulation is ensured by requiring compensation $\Pi(k)$. This scheme differs from feature (c) of actual IIAs 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, and thus assumes a Vickrey-Clarke-Groves type of mechanism.

5.2 A relative performance mechanism

The above compensation mechanisms all have their merits, either in terms of simplicity (carve-out policies), incentive compatibility (linear compensation as in (13)), or non-reliance on efficient investments (as in (14)). But they also have their shortcomings, either in terms of the possibility to reach an efficient outcome or when it comes to practical implementation. We show in Appendix A.11 that it is possible to implement the fully efficient outcome and simultaneously avoid drawbacks of the earlier models. Specifically, we let compensation be based on firms' *relative performance*. Such compensation could be relevant for cases where the same regulatory intervention affects multiple firms, so that several firms are potentially eligible for compensation. The claim below is made more precise and proved in Appendix A.11:

Observation 7 *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 3.*

The specific efficient compensation scheme we identify (expression (A.15) in the Appendix) differs from (13) by being based on the investments that firms have actually made, rather than on counterfactual earnings in an efficient outcome. Furthermore, the host country never overcompensates the firms. The rule differs from (14) by not relying on third-party participation. Instead, it 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 on the performance of similar firms, each firm is compensated for its operating profit in equilibrium.

The suggested compensation scheme has several other attractive features. First, it can implement the fully efficient outcome without information regarding the shock θ. Second, our mechanism does not depend on host country political incentives being observable. Third, it is independent of the firm's own profit, and no firm therefore has any unilateral incentive to misreport it.[33] Finally, the efficiency of the relative performance mechanism does not depend on firms being identical. But the mechanism does fail to achieve the first-best efficient solution if firms are very dissimilar, or if the industry consists of a single firm. Also, like the other schemes considered in this Section, it is incompatible with basic properties of actual IIAs, especially by requiring compensation to each firm to diverge from its foregone operating profits.

5.3 Asymmetric information regarding regulatory shocks

We have thus far for the most part assumed ex post verifiability of the regulatory shocks. But certain aspects of investment agreements seem better understood assuming that host countries have private information. For instance, Section 4.7 suggested that the stricter compensation rules for direct than for regulatory expropriations are best understood as addressing asymmetric information problems.

In our main setting, if θ is observed by the host country government only, it is no longer possible to implement an agreement that awards compensation below a threshold θ' but no compensation above it, because the host country would simply claim $\theta > \theta'$ for all realizations $\theta > \theta^0$ to be able to regulate without paying compensation. Incentive compatibility thus generally requires host countries to compensate firms for regulatory interventions (the proof is provided in Appendix A.12):

Proposition 12 *Assume that only the host country observes the realization of θ. For any agreement satisfying features (a)-(c) of IIAs, for which compensation at most equals foregone operating profit $\Pi(k)$, there exists an alternative, incentive compatible, agreement satisfying the same restrictions that:*
(i) features the compensation function

$$T(k,\theta) \equiv \begin{cases} \Pi(k) & \text{if } \theta \leq \theta' \\ \max\{-V(k,\theta');0\} & \text{if } \theta > \theta'; \end{cases}$$

(ii) yields weakly higher expected domestic welfare for the host country, and weakly higher foreign industry profits, than the initial agreement.

As in the full information case, the optimal agreement under asymmetric information requests compensation $\Pi(k)$ for $\theta \leq \theta'$. Whereas the host country will not be requested under complete information to compensate firms subsequent to regulation when $\theta > \theta'$, compensation is required in the asymmetric information case to achieve incentive compatibility. There will therefore be compensation payments in equilibrium regardless of whether $\theta' \gtrless \theta^E$. The optimal compensation scheme in Proposition 12 violates feature (e) of IIAs in two closely related respects, for $\theta > \theta'$. First,

[33] Myerson and Satterthwaite (1983) derive an optimal compensation mechanism with two-sided asymmetric information and a single firm. Their scheme requests payments even if there is no regulation and therefore violates our feature (b) of actual IIAs. They also consider the case where an arbitrator breaks the payment balance, in contradiction to feature (c).

compensation is not based on foregone operating profits, but on the value to the host country of shutting down production. Second, investors will not receive full compensation, since $-V(k,\theta') < \Pi(k)$.[34]

6 Concluding remarks

There are close to 2 700 international investment agreements in force, and protection of foreign direct investment has become a core issue in the policy debate. But the economic literature hardly sheds any light on their implications. This paper seeks to contribute to filling this void by examining agreements that protect investors against regulatory expropriations. It generates a large number of results that in our view yield insights into the functioning of investment agreements and the validity of core arguments in the policy debate.

Despite the simple legal structure of compensation payments in IIAs, they are still sufficient to implement efficient outcomes in a robust set of circumstances. This happens, for instance, when there are positive expected marginal externalities from the investment, and agreements are formed between fairly symmetric economies. Our findings also suggest that optimal agreements generally do not cause regulatory chill (underregulation) from a joint welfare perspective, provided decision makers are sufficiently concerned with maximizing domestic welfare. Furthermore, non-discrimination clauses are generally insufficient to solve investment and regulatory distortions. On the other hand, agreements can have pronounced distributional implications, benefitting foreign investors at the expense of the rest of society.

There are many potentially important aspects of IIAs that we have not addressed; we point to three such aspects here. First, we do not discuss the ISDS mechanism itself. Very little is known about how an ISDS system differs from one that only allows state-to-state litigation with regard to enforcement, and the willingness of countries to make investment protection undertakings. Second, it has become increasingly common to include investment protection in trade agreements. Complementarities between trade and investment undertakings can emanate for instance from global value chains, or reflect an exchange of concessions in the investment and trade areas (see e.g. Horstmann et al., 2005, and Maggi, 2016). However, the precise form of interaction between investment and trade undertakings remains to be identified. Third, we have considered the formation of one single investment agreement. It is often maintained that North-South agreements reflect a "race-to-the-bottom" for developing countries to attract foreign investment. It would therefore be desirable to allow for interrelated negotiations over investment agreements.

References

Aisbett, Emma, Larry Karp and Carol McAusland (2010a). "Police Powers, Regulatory Takings

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

and the Efficient Compensation of Domestic and Foreign Investors." *The Economic Record* 86(274), September, 367–383.

Aisbett, Emma, Larry Karp and Carol McAusland (2010b). "Compensation for Indirect Expropriation in International Investment Agreements: Implications of National Treatment and Rights to Invest." *Journal of Globalization and Development* 1(2), Article 6.

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

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

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

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

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

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

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

Hermalin, Benjamin E. (1995). "An Economic Analysis of Takings." *Journal of Law, Economics & Organization* 11(1), 64-86.

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

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

Horstmann, Ignatius J., Markusen, James R. and Jack Robles (2005). "Issue Linking in Trade Negotiations: Ricardo Revisited or No Pain No Gain." *Review of International Economics*, 13(2), 185–204.

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

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

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

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

Kohler, Wilhelm and Frank Stähler (2016). "The Economics of Investor Protection: ISDS versus National Treatment." CESifo Working Paper 5766.

Konrad, Kai A. (2016). "Large Investors and Permissive Regulation: Why Environmentalists May Dislike Investor-State Dispute Settlement." *European Economic Review 98,* 341-353.

Maggi, Giovanni (2016). "Issue Linkage." In Bagwell, Kyle and Robert W. Staiger (eds), *Handbook of Commercial Policy,* Vol. 1, Part B, 513-564. Amsterdam: North-Holland.

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

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

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

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

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

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

Saggi, Kamal and Nese Sara (2008). "National Treatment at the WTO: The Roles of Product and Country Heterogeneity." *International Economic Review* 49(4), 1365-1394.

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

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.

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

Transnational Institute (2017). "Why Did Equador Terminate All Its Bilateral Investment Treaties?" www.tni.org/en/work-area/trade-investment, May 9.

A Appendix

A.1 Proof of Lemma 1

Part (a): The joint surplus function of the host country and the firm, when defined in terms of k, is

$$\Omega(k) \equiv \int_{\underline{\theta}}^{\Theta^J(k)} [V(k,\theta) + \Pi(k)]dF(\theta) - R(k) \tag{A.1}$$

To see that this function is strictly quasi-concave, evaluate the second-order derivative at a solution $\Omega_k(k^J) = 0$, with $\theta^J = \Theta^J(k^J)$:

$$\frac{\Omega_{kk}(k^J)}{\Pi_{kk}(k^J)F(\theta^J) - R_{kk}(k^J) + X_k(k^J,\theta^J)} = 1 - \Theta_k^J(k^J)K_\theta^J(\theta^J).$$

The right-hand side is strictly positive by assumption (7), and since the denominator on the left-hand side is strictly negative, $\Omega_{kk}(k^J) < 0$ follows. As every solution to the first-order condition is a local maximum, $\Omega(k)$ is strictly quasi-concave with a unique global optimum k^J. Condition (8) is equivalent to $\Omega_k(k^0) > 0$, which implies $k^J > k^0$.

Part (b): The joint surplus function can alternatively be defined in terms of the regulatory threshold $\hat{\theta}$:

$$\Upsilon(\hat{\theta}) \equiv \int_{\underline{\theta}}^{\hat{\theta}} [V(K^J(\hat{\theta}),\theta) + \Pi(K^J(\hat{\theta}))]dF(\theta) - R(K^J(\hat{\theta})). \tag{A.2}$$

To see that this function is strictly quasi-concave, invoke (4) to get the first row of

$$\begin{aligned} \frac{\Upsilon_\theta(\hat{\theta})}{f(\hat{\theta})} &= V(K^J(\hat{\theta}),\hat{\theta}) + \Pi(K^J(\hat{\theta})) \\ &= V(K^J(\hat{\theta}),\hat{\theta}) - V(K^J(\hat{\theta}),\Theta^J(K^J(\hat{\theta}))). \end{aligned}$$

The second row follows from the definition of $\Theta^J(k)$. Assumption (7) and $V_\theta(k,\theta) < 0$ imply $\Upsilon_\theta(\hat{\theta}) > (<) 0$ for all $\hat{\theta} < (>) \theta^J$. Hence, $\Upsilon(\hat{\theta})$ is strictly quasi-concave with a unique maximum θ^J. The optimal investment k^J then is uniquely defined by $k^J = K^J(\theta^J)$. By strict quasi-concavity of $\Upsilon(\hat{\theta})$, $\theta^J > \theta^0$ if and only if $\Upsilon_\theta(\theta^0) > 0$, which is equivalent to (9). ■

A.2 Proof of Proposition 1

The stability condition (11) implies that the total expected welfare is defined by $\tilde{W}(\theta') = \Omega(K(\theta'))$ for investment protection levels $\theta' \geq \theta^E$, with $\Omega(k)$ as defined in (A.1). Strict quasi-concavity of $\Omega(k)$ implies that $\tilde{W}(\theta')$ has a unique maximum if constrained to be in $[\theta^E, \bar{\theta}]$. Assuming that full investment protection leads to overinvestment, $K(\bar{\theta}) \geq k^J$, it follows that $\tilde{W}_\theta(\bar{\theta}) = \Omega_k(K(\bar{\theta}))K_\theta(\bar{\theta}) \leq 0$.

Observe also that

$$\tilde{W}_\theta(\theta^E) = \Omega_k(k^E)K_\theta(\theta^E) = X(k^E, \theta^E)K_\theta(\theta^E)$$

is non-negative if $X(k^E, \theta^E) \geq 0$. In this case, there exists a level of investment protection $\theta^N \in [\theta^E, \bar{\theta}]$ satisfying $\tilde{W}_\theta(\theta^N) = \Omega_k(K(\theta^N))K_\theta(\theta^N) = 0$. Since $\Omega(k)$ has a unique maximum k^J, $K(\theta^N) = k^J$. The threshold for regulation is given by $\min\{\theta^N; \Theta^J(K(\theta^N))\} = \Theta^J(K(\theta^N)) = \Theta^J(k^J) = \theta^J$. Hence, θ^N implements the unconstrained joint optimum (k^J, θ^J) as a Nash equilibrium if $X(k^E, \theta^E) \geq 0$ and $K(\bar{\theta}) > k^J$.

Assume next that $X(k^E, \theta^E) < 0$, so that $\tilde{W}_\theta(\theta^E) < 0$. The optimal level of investment protection in $[\theta^E, \bar{\theta}]$ is then $\theta' = \theta^E$. By differentiability of the welfare function $\tilde{W}(\theta')$, $\tilde{W}_\theta(\theta^E) < 0$ implies $\theta^N < \theta^E$. The optimal level of investment protection θ^N then satisfies

$$[V(k^N, \theta^N) + \Pi(k^N)]f(\theta^N) + X(k^N, \theta^N)K_\theta(\theta^N) \leq 0,$$

where $k^N = K(\theta^N)$. By stability condition (11), $\theta^N < \theta^E$ implies $\theta^N < \Theta^J(k^N)$. Furthermore, $V(k^N, \theta^N) + \Pi(k^N) = V(k^N, \theta^N) - V(k^N, \Theta^J(k^N)) > 0$, so $X(k^N, \theta^N) < 0$. It then follows from Observation 2 that $k^N > K^J(\theta^N)$.■

A.3 Proof of Proposition 3 and Corollary 2

We prove statements made in Section 4 relating to the threshold property in Section 4.2 by way of a framework that in several respects is more general than the one employed in the main text. First, we assume that compensation schemes satisfy standard features (a)-(c) of IIAs, but we do not impose (d) and (e). Second, we make more general assumptions concerning industry structure, for instance by allowing firms to be heterogenous. Third, we employ two different assumptions regarding strategic investor behavior. We establish that any optimal investment agreement can be characterized in terms of a threshold that yields domestic, but never global regulatory chill both when investors are regulation strategic (Section A.3.2) and regulation takers (Section A.3.3). Proposition 3 in the main text then is a special case of Section A.3.3.

A.3.1 A generalized model

There are two countries, $i \neq j = 1, 2$, and an industry with $H \geq 1$ firms, indexed $h = 1, 2..., H$. Assume that each firm h invests k_{hi} in country i, so that $\mathbf{k}_h = (k_{h1}, k_{h2})$ is the firm's investment portfolio. Let $\mathbf{k}_{-hi} = (k_{1i}, .., k_{(h-1)i}, k_{(h+1)i}, .., k_{Hi})$ be the investment profile of all firms in country i other than h, and denote by $\mathbf{k}_i = (k_{hi}, \mathbf{k}_{-hi})$ the full portfolio of investments in country i. The firms make their investment decisions simultaneously and independently to maximize their respective profits, but we do not make any assumptions about the nature of strategic interaction in the investment stage nor in the product market. Let $\Pi^{hi}(\mathbf{k}_i) \geq 0$ be the reduced form operating profit of firm h of its facilities in country i, and assume that it is independent of whether country j is regulated or not. Obviously, $\Pi^{hi}(\mathbf{k}_i) = 0$ if firm h does not have any facilities in country i. Denote by $R^h(\mathbf{k}_h) \geq 0$ firm h's rental cost of capital, which is strictly positive if $k_{h1} > 0$ and/or $k_{h2} > 0$.

The domestic welfare is $V^i(\mathbf{k}_i, \theta_i)$ if there is production, and zero if there is regulation. The domestic welfare depends on investment $\mathbf{k}_i$ in the home country and on the 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 $k_{hi} > 0$ for at least one firm and a strict inequality that investments are strictly positive for all firms). Assume that $V^i(\mathbf{k}_i, \theta_i)$ and the profit functions are continuous in $\mathbf{k}_i$.

Assume that θ_i is continuously distributed on $[\underline{\theta}_i, \bar{\theta}_i]$ with cumulative distribution function $F^i(\theta_i)$ and density $f^i(\theta_i)$. 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 production in all firms in the industry in the host country.

Consider now the ex post optimal regulatory choice of country i. Assume that country i attaches the weight $\gamma_{hi} \in [0, 1]$ to the profit of firm h in its decision whether to regulate, and let $\boldsymbol{\gamma}_i = (\gamma_{1i}, .., \gamma_{hi}, .., \gamma_{Hi})$. Denote by

$$\Delta^i(\mathbf{k}_i, \theta_i, \boldsymbol{\gamma}_i) \equiv V^i(\mathbf{k}_i, \theta_i) + \textstyle\sum_{h=1}^{H} \gamma_{hi} \Pi^{hi}(\mathbf{k}_i)$$

the net benefit to country i of allowing production. Let $\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \equiv \underline{\theta}_i$ if $\Delta^i(\mathbf{k}_i, \underline{\theta}_i, \boldsymbol{\gamma}_i) \leq 0$, $\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \equiv \bar{\theta}_i$ if $\Delta^i(\mathbf{k}_i, \bar{\theta}_i, \boldsymbol{\gamma}_i) \geq 0$, and let $\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ be the implicit solution to $\Delta^i(\mathbf{k}_i, \Theta^i, \boldsymbol{\gamma}_i) \equiv 0$ in intermediate cases. We assume that the host country allows production if indifferent. Country i will then allow production if and only if $\theta_i \leq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. Country i's regulatory decision is independent of country j's actions due to the separability of the industries (the interrelationship that stems from the investment cost does not affect regulatory decisions). Define the threshold $\Theta^{iJ}(\mathbf{k}_i) \equiv \Theta^i(\mathbf{k}_i, \mathbf{1}) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. This is the cut-off below which it is ex post jointly optimal to allow production in country i.

An *international investment agreement* (IIA) is a vector $\mathbf{T}^i = (T^{1i}, .., T^{hi}, .., T^{Hi})$ of compensation rules for each country, where $T^{hi}(\mathbf{k}_i, \theta_i) \geq 0$ specifies the compensation from the host country to firm h in case of regulation in country i. Notice that the compensation rule only depends on domestic factors; countries thus never compensate for regulation abroad. Also, it assumes that $\mathbf{k}_i$ and θ_i are verifiable. Observe that $\mathbf{T}^i$ satisfies contract features (a)-(c), we have not imposed (d) and (e).

The timing of the interaction is as follows:

1. The two countries jointly commit to an IIA with compensation rules $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$;
2. Firms invest $\mathbf{k}$;
3. Country i observes the shock θ_i and decides whether to regulate.

(a) If country i does not intervene, then product market competition ensues in country i;
(b) If country i regulates, then the agreement pays compensation according to $\mathbf{T}^i$.

A.3.2 Regulation-strategic investors

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

$$\begin{aligned} M^i(\mathbf{k}_i) \equiv &\quad \{\theta_i : \Delta^i(\mathbf{k}_i, \theta_i, \boldsymbol{\gamma}_i) + \textstyle\sum_{h=1}^{H}(1-\gamma_{hi})T^{hi}(\mathbf{k}_i, \theta_i)\} \geq 0\}, \\ M^{ir}(\mathbf{k}_i) \equiv &\quad \{\theta_i \notin M^i(\mathbf{k}_i)\}. \end{aligned} \tag{A.3}$$

$M^i(\mathbf{k}_i)$ and $M^{ir}(\mathbf{k}_i)$ also depend on $\boldsymbol{\gamma}_i$, but we subsume $\boldsymbol{\gamma}_i$ for notational simplicity. The second component of the SPE is the investment profile $\hat{\mathbf{k}}_h = (\hat{k}_{h1}, \hat{k}_{h2})$, which for $h = 1, 2..., H$ is given by

$$\begin{aligned} \hat{\mathbf{k}}_h \in &\quad \underset{\mathbf{k}_h \in \mathbb{R}_+^2}{\arg\max} \{\textstyle\sum_{i=1,2}[\Pi^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi})\int_{M^i(k_{hi},\hat{\mathbf{k}}_{-hi})} dF^i(\theta_i) \\ + &\quad \textstyle\int_{M^{ir}(k_{hi},\hat{\mathbf{k}}_{-hi})} T^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}, \theta_i) dF^i(\theta_i)] - R^h(\mathbf{k}_h)\}. \end{aligned} \tag{A.4}$$

where $\hat{\mathbf{k}}_{-hi} = (\hat{k}_{1i}, .., \hat{k}_{(h-1)i}, \hat{k}_{(h+1)i}, .., \hat{k}_{Hi})$ is the equilibrium investment profile for all firms except h in country i.

Equilibrium expected profit and host country welfare. Let $\hat{m}_i \equiv M^i(\hat{\mathbf{k}}_i)$ be the subset of shocks for which country i allows production in equilibrium, and let $\hat{m}_{ir} \equiv M^{ir}(\hat{\mathbf{k}}_i)$ be the events with regulation. Then

$$\tilde{\Pi}^{hi}(\mathbf{T}) \equiv \Pi^{hi}(\hat{\mathbf{k}}_i)\textstyle\int_{\hat{m}_i} dF^i(\theta_i) + \int_{\hat{m}_{ir}} T^{hi}(\hat{\mathbf{k}}_i, \theta_i) dF^i(\theta_i) \tag{A.5}$$

is the equilibrium expected operating profit of firm h in market i, and $\tilde{\Pi}^h(\mathbf{T}) \equiv \tilde{\Pi}^{h1}(\mathbf{T}) + \tilde{\Pi}^{h2}(\mathbf{T})$ the total expected profit excluding capital costs $R^h(\hat{\mathbf{k}}_h)$. The expected surplus of the government in host country i equals

$$\begin{aligned} \tilde{V}^i(\mathbf{T}, \boldsymbol{\gamma}_i) \quad \equiv \quad & \textstyle\int_{\hat{m}_i}(V^i(\hat{\mathbf{k}}_i, \theta_i) + \sum_{h=1}^{H} \gamma_{hi}\Pi^{hi}(\hat{\mathbf{k}}_i)) dF^i(\theta_i) \\ & -\textstyle\int_{\hat{m}_{ir}} \sum_{h=1}^{H}(1-\gamma_{hi}) T^{hi}(\hat{\mathbf{k}}_i, \theta_i) dF^i(\theta_i) \\ & +\textstyle\sum_{h=1}^{H} \gamma_{hi}[\Pi^{hj}(\hat{\mathbf{k}}_j)\int_{\hat{m}_j} dF^j(\theta_j) \\ & +\textstyle\int_{\hat{m}_{jr}} T^{hj}(\hat{\mathbf{k}}_j, \theta_j) dF^j(\theta_j)] - \sum_{h=1}^{H} \gamma_{hi} R^h(\hat{\mathbf{k}}_h). \end{aligned}$$

Let $\hat{\theta}_i^J \equiv \Theta^{iJ}(\hat{\mathbf{k}}_i)$ be the ex post jointly optimal level of regulation given the equilibrium investment $\hat{\mathbf{k}}_i$, so that $V^i(\hat{\mathbf{k}}_i, \hat{\theta}_i^J) + \sum_{h=1}^{H} \Pi^{hi}(\hat{\mathbf{k}}_i) = 0$. Define the expected operating surplus in country i as

$$\tilde{Z}^i(\mathbf{T}) \equiv \textstyle\int_{\hat{m}_i}(V^i(\hat{\mathbf{k}}_i, \theta_i) - V^i(\hat{\mathbf{k}}_i, \hat{\theta}_i^J)) dF^i(\theta_i). \tag{A.6}$$

We can then write the expected surplus of the government in host country i more compactly as

$$\tilde{V}^i(\mathbf{T},\boldsymbol{\gamma}_i)=\tilde{Z}^i(\mathbf{T})+\textstyle\sum_{h=1}^{H}[\gamma_{hi}(\tilde{\Pi}^h(\hat{\mathbf{T}})-R^h(\hat{\mathbf{k}}_h))-\tilde{\Pi}^{hi}(\hat{\mathbf{T}})]. \tag{A.7}$$

Proposition A.1 *Assume that all firms account for the effect of their investment on regulation. For any feasible international investment agreement* $\mathbf{T}$ *satisfying (a)-(c), there exists an alternative agreement* $\hat{\mathbf{T}}$ *characterized by a threshold* $\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)\in[\Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i),\Theta^{iJ}(\mathbf{k}_i)]$ *for host country* i*. This alternative agreement* $\hat{\mathbf{T}}$ *satisfies (a)-(c) and yields:*
(i) regulation in host country i *if and only if* $\theta_i>\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)$*;*
(ii) the same equilibrium investment $\hat{\mathbf{k}}$ *as the initial agreement* $\mathbf{T}$*;*
(iii) weakly higher expected welfare and industry profits than the initial agreement $\mathbf{T}$*.*

Proof: We first use the threshold function $\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)$ (defined below) to create four partitions of $[\underline{\theta}_i,\bar{\theta}_i]$:

$$\begin{aligned}
A^i(\mathbf{k}_i) &\equiv \{\theta_i\in M^i(\mathbf{k}_i)\cap[\underline{\theta}_i,\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)]\}\\
A^{ir}(\mathbf{k}_i) &\equiv \{\theta_i\in M^{ir}(\mathbf{k}_i)\cap[\underline{\theta}_i,\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)]\}\\
B^i(\mathbf{k}_i) &\equiv \{\theta_i\in M^i(\mathbf{k}_i)\cap(\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i),\bar{\theta}_i]\}\\
B^{ir}(\mathbf{k}_i) &\equiv \{\theta_i\in M^{ir}(\mathbf{k}_i)\cap(\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i),\bar{\theta}_i]\}
\end{aligned}$$

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

An alternative investment agreement. Let the agreement $\hat{\mathbf{T}}=(\hat{\mathbf{T}}^1,\hat{\mathbf{T}}^2)$ be characterized by a threshold $\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)$ for each country given by

$$F^i(\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i))\equiv\min\{\textstyle\int_{M^i(\mathbf{k}_i)}dF^i(\theta_i);F^i(\Theta^{iJ}(\mathbf{k}_i))\} \tag{A.8}$$

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

$$\hat{T}^{hi}(\mathbf{k}_i,\theta_i)=\begin{cases}\Pi^{hi}(\mathbf{k}_i) & \theta_i\in A^i(\mathbf{k}_i)\cup A^{ir}(\mathbf{k}_i)\\ \tilde{T}^{hi}(\mathbf{k}_i) & \theta_i\in B^i(\mathbf{k}_i)\\ T^{hi}(\mathbf{k}_i,\theta_i) & \theta_i\in B^{ir}(\mathbf{k}_i)\end{cases}, \tag{A.9}$$

and where

$$\begin{aligned}\tilde{T}^{hi}(\mathbf{k}_i)\equiv &\quad \tfrac{1}{\int_{B^i(\mathbf{k}_i)}dF^i(\tilde{\theta}_i)}[\textstyle\int_{A^{ir}(\mathbf{k}_i)}T^{hi}(\mathbf{k}_i,\tilde{\theta}_i)dF^i(\tilde{\theta}_i)\\ &+\max\{\textstyle\int_{M^i(\mathbf{k}_i)}dF^i(\tilde{\theta}_i)-F^i(\Theta^{iJ}(\mathbf{k}_i));0\}\Pi^{hi}(\mathbf{k}_i)]\end{aligned} \tag{A.10}$$

if $\int_{B^i(\mathbf{k}_i)} dF^i(\tilde{\theta}_i) > 0$, and where

$$T^{hi}(\mathbf{k}_i, \theta_i) \equiv \begin{cases} \Pi^{hi}(\mathbf{k}_i) & \text{if } \theta_i \le \hat{\Theta}^i(\mathbf{k}, \boldsymbol{\gamma}_i) \\ \tilde{\Lambda}^i(\mathbf{k}_i,\theta_i)\Pi^{hi}(\mathbf{k}_i) + \hat{\Lambda}^i(\mathbf{k}_i,\theta_i)\hat{T}^{hi}(\mathbf{k}_i,\theta_i) + \int_{\underline{\theta}_i}^{\bar{\theta}_i} \hat{T}^i(k_i,\tilde{\theta}_i) d\Lambda^i(\mathbf{k}_i,\tilde{\theta}_i) & \text{if } \theta_i > \hat{\Theta}^i(\mathbf{k}, \boldsymbol{\gamma}_i), \end{cases}$$

with

$$\begin{aligned} \tilde{\Lambda}^i(\mathbf{k}_i,\theta_i) &\ge 0 \\ \hat{\Lambda}^i(\mathbf{k}_i,\theta_i) &\ge 0 \\ \Lambda^i(\mathbf{k}_i,\theta_i) &\ge 0 \\ \tilde{\Lambda}^i(\mathbf{k}_i,\theta_i) + \hat{\Lambda}^i(\mathbf{k}_i,\theta_i) + \int_{\underline{\theta}_i}^{\bar{\theta}_i} d\Lambda^i(\mathbf{k}_i,\tilde{\theta}_i) &= 1. \end{aligned}$$

This alternative agreement builds on the payments under the original compensation scheme and the operating profits of regulated firms. $\hat{\Theta}^i$ therefore depends on $\boldsymbol{\gamma}_i$ since $M^i(\mathbf{k}_i)$ depends on $\boldsymbol{\gamma}_i$.

Establishing $\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i) \in [\Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i), \Theta^{iJ}(\mathbf{k}_i)]$. The inequality $\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i) \le \Theta^{iJ}(\mathbf{k}_i)$ follows directly from (A.8). Furthermore, $\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i) \ge \Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)$ trivially holds if $\Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i) = \underline{\theta}_i$. To establish $\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i) \ge \Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)$ for $\Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i) > \underline{\theta}_i$, note that if $F^i(\Theta^{iJ}(\mathbf{k}_i)) \le \int_{M^i(\mathbf{k}_i)} dF^i(\theta_i)$, then $F^i(\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)) = F^i(\Theta^{iJ}(\mathbf{k}_i)) \ge F^i(\Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i))$, where the inequality follows from $\Theta^{iJ}(\mathbf{k}_i) \ge \Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)$. Assume finally that $\int_{M^i(\mathbf{k}_i)} dF^i(\theta_i) < F^i(\Theta^{iJ}(\mathbf{k}_i))$. The assumption that $V^i(\mathbf{k}_i,\theta_i)$ is strictly decreasing in θ_i and $T^{hi}(\mathbf{k}_i,\theta_i) \ge 0$ jointly imply that

$$\Delta^i(\mathbf{k}_i,\theta_i,\boldsymbol{\gamma}_i) + \textstyle\sum_{h=1}^H (1-\gamma_{hi}) T^{hi}(\mathbf{k}_i,\theta_i)$$

is strictly positive for all $\theta_i < \Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)$ in the initial agreement. Hence, $[\underline{\theta}_i, \Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)) \subset M^i(\mathbf{k}_i)$ and therefore $F^i(\hat{\Theta}^i(\mathbf{k}_i,\boldsymbol{\gamma}_i)) = \int_{M^i(\mathbf{k}_i)} dF^i(\theta_i) \ge F^i(\Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i))$.

Country i regulates under agreement $\hat{\mathbf{T}}$ iff $\theta_i > \hat{\Theta}^i(\mathbf{k},\boldsymbol{\gamma}_i)$. 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) $\theta_i \in A^i(\mathbf{k}_i) \cup A^{ir}(\mathbf{k}_i) = [\underline{\theta}_i, \hat{\Theta}^i(\mathbf{k},\boldsymbol{\gamma}_i)]$. By construction of the agreement, the net benefit of allowing production is non-negative for all $\theta_i \le \hat{\Theta}^i(\mathbf{k},\boldsymbol{\gamma}_i) \le \Theta^{iJ}(\mathbf{k}_i)$ because in this case

$$\Delta^i(\mathbf{k}_i,\theta_i,\boldsymbol{\gamma}_i) + \textstyle\sum_{h=1}^H (1-\gamma_{hi}) T^{hi}(\mathbf{k}_i,\theta_i) = V^i(\mathbf{k}_i,\theta_i) + \sum_{h=1}^H \Pi^{hi}(\mathbf{k}_i) \ge 0.$$

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

$$\int_{B^i(\mathbf{k}_i)} dF^i(\tilde{\theta}_i) \equiv \int_{A^{ir}(\mathbf{k}_i)} dF^i(\tilde{\theta}_i) + \max\{\textstyle\int_{M^i(\mathbf{k}_i)} dF^i(\tilde{\theta}_i) - F(\Theta^{iJ}(\mathbf{k}_i)); 0\}. \tag{A.11}$$

Use $\tilde{T}^{hi}(\mathbf{k}_i)$ defined in (A.10) and (A.11) to decompose the net benefit of allowing production in country i as follows:

$$\begin{aligned}
&\Delta^i(\mathbf{k}_i,\theta_i,\boldsymbol{\gamma}_i)+\textstyle\sum_{h=1}^H(1-\gamma_{hi})\tilde{T}^{hi}(\mathbf{k}_i)\\
=\;&\frac{\int_{A^{ir}(\mathbf{k}_i)}[V^i(\mathbf{k}_i,\theta_i)-V^i(\mathbf{k}_i,\tilde{\theta}_i)]dF^i(\tilde{\theta}_i)}{\int_{B^i(\mathbf{k}_i)}dF^i(\tilde{\theta}_i)}\\
&+\frac{\int_{A^{ir}(\mathbf{k}_i,\hat{\mathbf{T}}^i)}[\Delta^i(\mathbf{k}_i,\tilde{\theta}_i,\boldsymbol{\gamma}_i)+\sum_{h=1}^H(1-\gamma_{hi})T^{hi}(\mathbf{k}_i,\tilde{\theta}_i)]dF^i(\tilde{\theta}_i)}{\int_{B^i(\mathbf{k}_i)}dF^i(\tilde{\theta}_i)}\\
&+\frac{[V^i(\mathbf{k}_i,\theta_i)-V^i(\mathbf{k}_i,\Theta^{iJ}(\mathbf{k}_i))]\max\{\int_{M^i(\mathbf{k}_i)}dF^i(\tilde{\theta}_i)-F^i(\Theta^{iJ}(\mathbf{k}_i));0\}}{\int_{B^i(\mathbf{k}_i)}dF^i(\tilde{\theta}_i)}
\end{aligned}$$

Assume first that $\int_{A^{ir}(\mathbf{k}_i)}dF^i(\theta_i)>0$. In this case, the term on the second row is strictly negative because $V^i_\theta<0$ and $\theta_i>\hat{\Theta}^i(\mathbf{k},\boldsymbol{\gamma}_i)\geq\tilde{\theta}_i$ for all $\theta_i\in B^i(\mathbf{k}_i)$ and $\tilde{\theta}_i\in A^{ir}(\mathbf{k}_i)$. The term on the third row is strictly negative because regulation is optimal under contract $\mathbf{T}$ for all $\tilde{\theta}_i\in A^{ir}(\mathbf{k}_i)$. The term on the fourth row is zero if $\int_{M^i(\mathbf{k}_i)}dF^i(\tilde{\theta}_i)\leq F^i(\Theta^{iJ}(\mathbf{k}_i))$ and strictly negative otherwise because then $\theta_i>\hat{\Theta}^i(\mathbf{k},\boldsymbol{\gamma}_i)=\Theta^{iJ}(\mathbf{k}_i)$ for all $\theta_i\in B^i(\mathbf{k}_i)$. The terms on the second and third row vanish if $\int_{A^{ir}(\mathbf{k}_i)}dF^i(\theta_i)=0$. But then $\int_{M^i(\mathbf{k}_i)}dF^i(\tilde{\theta}_i)>F(\Theta^{iJ}(\mathbf{k}_i))$ by (A.11) so the third term is strictly negative.

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

Investments and expected profits are the same under both agreements. By way of the threshold $\hat{\Theta}^i(\mathbf{k},\boldsymbol{\gamma}_i)$ for regulation defined in (A.8) and the compensation rules (A.9)-(A.10), the expected operating profit of firm h active in country i under the modified agreement $\hat{\mathbf{T}}$ becomes

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

after simplifications. This is *exactly* the same expected operating profit as under the original agreement $\mathbf{T}$ 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 $\hat{\mathbf{T}}$.

It follows directly from the observation that operating profits and the equilibrium investments are the same under both agreements that $\tilde{\Pi}^{h1}(\hat{\mathbf{T}})=\tilde{\Pi}^{h1}(\mathbf{T})$, $\tilde{\Pi}^{h2}(\hat{\mathbf{T}})=\tilde{\Pi}^{h2}(\mathbf{T})$ and $\tilde{\Pi}^{h}(\hat{\mathbf{T}})=\tilde{\Pi}^{h}(\mathbf{T})$ for all h.

Expected host country surplus is weakly higher under agreement $\hat{\mathbf{T}}$. The equilibrium surplus of the government in host country i equals

$$\begin{aligned}\tilde{V}^i(\hat{\mathbf{T}},\boldsymbol{\gamma}_i) &\equiv \tilde{Z}^i(\hat{\mathbf{T}})+\textstyle\sum_{h=1}^H[\gamma_{hi}(\tilde{\Pi}^h(\hat{\mathbf{T}})-R^h(\hat{\mathbf{k}}_h))-\tilde{\Pi}^{hi}(\hat{\mathbf{T}})] \\ &= \tilde{Z}^i(\hat{\mathbf{T}})+\textstyle\sum_{h=1}^H[\gamma_{hi}(\tilde{\Pi}^h(\mathbf{T})-R^h(\hat{\mathbf{k}}_h))-\tilde{\Pi}^{hi}(\mathbf{T})]\end{aligned}$$

under agreement $\hat{\mathbf{T}}$, where

$$\tilde{Z}^i(\hat{\mathbf{T}})\equiv\textstyle\int_{-\infty}^{\hat{\theta}_i}(V^i(\hat{\mathbf{k}}_i,\theta_i)-V^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^J))dF^i(\theta_i),$$

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

$$\begin{aligned}\tilde{V}^i(\hat{\mathbf{T}},\boldsymbol{\gamma}_i)-\tilde{V}^i(\mathbf{T},\boldsymbol{\gamma}_i) &= \tilde{Z}^i(\hat{\mathbf{T}})-\tilde{Z}^i(\mathbf{T}) \\ &= \textstyle\int_{\underline{\theta}_i}^{\hat{\theta}_i}(V^i(\hat{\mathbf{k}}_i,\theta_i)-V(\hat{\mathbf{k}}_i,\hat{\theta}_i^J))dF^i(\theta_i) \\ &\quad -\textstyle\int_{\hat{a}_i}(V^i(\hat{\mathbf{k}}_i,\theta_i)-V^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^J))dF^i(\theta_i) \\ &\quad -\textstyle\int_{\hat{b}_i}(V^i(\hat{\mathbf{k}}_i,\theta_i)-V^i(\hat{\mathbf{k}}_i,\hat{\theta}_i^J))dF^i(\theta_i),\end{aligned}$$

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

$$\begin{aligned}\tilde{Z}^i(\hat{\mathbf{T}})-\tilde{Z}^i(\mathbf{T}) &= \textstyle\int_{\hat{a}_{ir}}(V^i(\hat{\mathbf{k}}_i,\theta_i)-V^i(\hat{\mathbf{k}}_i,\hat{\theta}_i))dF^i(\theta_i) \\ &\quad +\textstyle\int_{\hat{b}_i}(V^i(\hat{\mathbf{k}}_i,\hat{\theta}_i)-V^i(\hat{\mathbf{k}}_i,\theta_i))dF^i(\theta_i) \\ &\quad +(V^i(\hat{\mathbf{k}}_i,\hat{\theta}_i)-V(\hat{\mathbf{k}}_i,\hat{\theta}_i^J))(F^i(\hat{\theta}_i)-\textstyle\int_{\hat{m}_i}dF^i(\theta_i))\end{aligned}$$

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

Remarks Our alternative compensation rule requests a compensation $\hat{T}^{hi}$ for firm h in country i that is a convex combination of that firm's operating profit Π^{hi} and the compensation T^{hi} in the original scheme, where the weights on the two components are country-specific and depend on $(\mathbf{k}_i,\theta_i)$, but are the same for all firms that have invested in country i. This structure implies that the modified scheme $\hat{\mathbf{T}}$ inherits a number of characteristics from the original scheme $\mathbf{T}$. First, compensation is non-negative because operating profit is non-negative and the original compensation is non-negative ($\Pi^{hi}\geq 0$ and $T^{hi}\geq 0$ imply $\hat{T}^{hi}\geq 0$). Second, it does not rely on excessive compensation (punitive damages) if this is not part of the original scheme ($T^{hi} \leq \Pi^{hi}$ implies $\hat{T}^{hi} \leq \Pi^{hi}$). 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. The statements in Proposition A.1 would thus hold also for stricter restrictions on IIAs than features (a)-(c). It also shows that linear compensation rules that incorporate both operating profits and incurred capital costs are superior to rules that compensate incurred capital costs only, as discussed in Section 5.

A.3.3 Regulation–taking investors

The above results are based on the assumption that firms take into account how their investments affect the probability of being regulated, in which case SPE is the appropriate equilibrium concept. We next assume that firms treat the probability of host country intervention as being exogenous to their own investment, in which case the Nash equilibrium is the appropriate equilibrium concept. Given the investment agreement $\mathbf{T}$, a Nash equilibrium 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 regulates and an investment profile $\hat{\mathbf{k}}_i$, such that allowing production and regulation are both ex post optimal given $\hat{\mathbf{k}}_i$ and the realization of the shock:

$$\begin{aligned} \hat{m}_i \equiv & \quad \{\theta_i : \Delta^i(\hat{\mathbf{k}}_i, \theta_i, \boldsymbol{\gamma}_i) + \textstyle\sum_{h=1}^{H}(1-\gamma_{hi})T^{hi}(\hat{\mathbf{k}}_i, \theta_i)\} \geq 0\}, \\ \hat{m}_{ir} \equiv & \quad \{\theta_i \notin \hat{m}_i\}, \end{aligned} \tag{A.12}$$

and $\hat{\mathbf{k}}_h = (\hat{k}_{h1}, \hat{k}_{h2})$ represents a profit maximizing investment portfolio of firm h given $\hat{\mathbf{k}}_{-hi}$, $\hat{m}_i$ and $\hat{m}_{ir}$:

$$\begin{aligned} \hat{\mathbf{k}}_h \in & \quad \arg\max_{\mathbf{k}_h \in \mathbb{R}^2_+} \{\textstyle\sum_{i=1,2}[\Pi^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi})\int_{\hat{m}_i} dF^i(\theta_i) \\ + & \quad \int_{\hat{m}_{ir}} T^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}, \theta_i) dF^i(\theta_i)] - R^h(\mathbf{k}_h)\}. \end{aligned} \tag{A.13}$$

Every SPE is contained in the set of Nash equilibria, so $(\hat{m}_i, \hat{m}_{ir})$ and $\hat{\mathbf{k}}_i$ as defined in (A.12)-(A.13), represent Nash equilibrium outcomes of the market game induced by IIA $\mathbf{T}$. The expected welfare $\tilde{V}^i(\mathbf{T}, \boldsymbol{\gamma}_i)$ of country i, and the operating profits $\tilde{\Pi}^{hi}(\mathbf{T})$ and $\tilde{\Pi}^h(\mathbf{T})$ of each firm h, are unaffected by this change in equilibrium concept.

Proposition A.2 *Proposition A.1 holds also under the assumption that each firm h treats regulation as exogenous to the own investment* $\mathbf{k}_h$.

Proof: For any initial agreement $\mathbf{T}$, define the modified agreement $\hat{\mathbf{T}}$ by (A.8)-(A.10). It is then optimal for country i to allow production if and only if $\theta_i \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \in [\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i), \Theta^{iJ}(\mathbf{k}_i)]$ for any realized investment profile $\mathbf{k}_i$, as was shown already in the proof of Proposition A.1. In particular, country i allows production for the investment profile $\hat{\mathbf{k}}_i$ if and only if $\theta_i \leq \hat{\theta}_i$, where

$$F^i(\hat{\theta}_i) \equiv \min\{\int_{\hat{m}_i} dF^i(\theta_i); F^i(\hat{\theta}_i^J)\}.$$

The expected operating profit of firm h active in country i becomes

$$\Pi^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi})F^i(\hat{\theta}_i) + \tilde{T}^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi})\int_{\hat{b}_i} dF^i(\theta_i) + \int_{\hat{b}_{ir}} T^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}, \theta_i)dF^i(\theta_i)$$
$$= \Pi^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi})\int_{\hat{m}_i} dF^i(\theta_i) + \int_{\hat{m}_{ir}} T^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}, \theta_i)dF^i(\theta_i)$$

under the modified agreement $\hat{\mathbf{T}}$, given the investment profile $\hat{\mathbf{k}}_{-hi}$ of all other firms active in country i and the expectation that $A^i(k_{hi}, \hat{\mathbf{k}}_{-hi}) = \hat{a}_i$, $A^{ir}(k_{hi}, \hat{\mathbf{k}}_{-hi}) = \hat{a}_{ir}$, and so forth. The expected operating profit is exactly the same as under the original agreement. Hence, the thresholds $(\hat{\theta}_1, \hat{\theta}_2)$ and investment profile $\hat{\mathbf{k}}$ can be implemented as a Nash Equilibrium by means of the compensation rules $\hat{\mathbf{T}} = (\hat{\mathbf{T}}^1, \hat{\mathbf{T}}^2)$. And since the surplus of the host country governments, all operating profits and investment are the same as before, it follows that the alternative agreement $\hat{\mathbf{T}}$ represents an expected improvement for all parties even under Nash implementation.■

Remarks Note that Proposition 3 in the main text is a special case of Proposition A.2 above, with one representative firm in each country investing only in FDI, and where $\gamma_{hi} = 0$ for the foreign firm h investing in country i.

Propositions A.1 and A.2 characterize the ex post optimal regulation for any θ_i, and the optimal compensation for $\theta_i \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. However, the propositions are silent about the optimal compensation for $\theta_i > \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ because the modified compensation scheme $\hat{\mathbf{T}} = (\hat{\mathbf{T}}^1, \hat{\mathbf{T}}^2)$ is defined relative to some initial and arbitrary compensation scheme $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$ in this case. We need to impose more structure on the compensation scheme to derive more specific features of optimal agreements.

A.4 Proof of Proposition 4

Proposition 4 obviously holds if $X(k^E, \theta^E) \geq 0$ and $k^J \leq K(\bar{\theta})$ because the compensation rule (10) can then implement the unconstrained joint optimum (k^J, θ^J). Consider therefore the case of $X(k^E, \theta^E) < 0$. Based on Propositions A.1 and A.2, and the assumption that the firm treats the probability of regulation as exogenous to its own investment, we can restrict attention to compensation rules T that in Nash equilibrium yield a regulatory threshold $\hat{\theta} \in [\Theta(\hat{k}), \Theta^J(\hat{k})]$, and with a corresponding equilibrium level of investment given by

$$F(\hat{\theta})\Pi_k(\hat{k}) + \int_{\hat{\theta}}^{\bar{\theta}} T_k(\hat{k}, \theta)dF(\theta) - R_k(\hat{k}) = 0$$

if $\hat{k} > 0$, where $T(\hat{k}, \theta)$ is differentiable almost everywhere by the assumption that $T(\hat{k}, \theta)$ is non-decreasing.

Now consider a specific class of rules that yields compensation for all regulation according to

$$\hat{T}(k, \tilde{\theta}, b) \equiv \max\{-V(\tilde{K}(\tilde{\theta}, b), \tilde{\theta}); 0\} + b[k - \tilde{K}(\tilde{\theta}, b)],$$

where $\tilde{K}(\tilde{\theta}, b)$ solves

$$F(\tilde{\theta})\Pi_k(\tilde{K}) + (1 - F(\tilde{\theta}))b - R_k(\tilde{K}) = 0$$

if $\tilde{K}(\tilde{\theta}, b) > 0$. By construction, the compensation function $\hat{T}(k, \tilde{\theta}, b)$ implements *any* regulatory threshold $\tilde{\theta} \geq \Theta(\tilde{K}(\tilde{\theta}, b))$ and the investment level $\tilde{K}(\tilde{\theta}, b)$ as a Nash equilibrium. We will show that it is optimal to set $b = 0$, that is, to use the compensation rule (10).

The expected joint surplus of the two contracting parties equals

$$\tilde{\Omega}(\tilde{\theta}, b) \equiv \int_{\underline{\theta}}^{\tilde{\theta}} [V(\tilde{K}(\tilde{\theta}, b), \theta) + \Pi(\tilde{K}(\tilde{\theta}, b))] dF(\theta) - R(\tilde{K}(\tilde{\theta}, b)).$$

In particular, the Nash equilibrium $(\hat{\theta}, \hat{k})$ induced by the non-decreasing compensation rule T above can be implemented by the modified rule by setting $\tilde{\theta} = \hat{\theta}$ and letting the compensation function be $\hat{T}(k, \hat{\theta}, \hat{b})$, where

$$(1 - F(\hat{\theta}))\hat{b} \equiv \int_{\hat{\theta}}^{\bar{\theta}} T_k(\hat{k}, \theta) dF(\theta) \geq 0$$

and the inequality follows by the assumption that $T(k, \theta)$ is non-decreasing in k. The expected joint surplus is the same under $\hat{T}(k, \hat{\theta}, \hat{b})$ as T and given by $\tilde{\Omega}(\hat{\theta}, \hat{b})$ because the compensation payment is lump-sum.

Assume that $(\tilde{\theta}, b)$ is chosen to maximize $\tilde{\Omega}(\tilde{\theta}, b)$ subject to $\tilde{\theta} \geq \Theta(\tilde{K}(\tilde{\theta}, b))$ and $b \geq 0$. The Lagrangian is

$$L(\tilde{\theta}, b) \equiv \tilde{\Omega}(\tilde{\theta}, b) + \lambda(\tilde{\theta} - \Theta(\tilde{K}(\tilde{\theta}, b))) + \chi b.$$

Let (θ^*, b^*) be a solution to this problem, with $k^* = \tilde{K}(\theta^*, b^*)$.

The optimal choice of b^*: If $\theta^* = \underline{\theta}$, then $\tilde{K}(\theta^*, b) > 0$ for all $b > 0$. By $V(k, \underline{\theta}) > 0$ for all $k > 0$, it follows that $\Theta(\tilde{K}(\underline{\theta}, b)) > \underline{\theta}$ and that $\underline{\theta}$ therefore is infeasible for $b > 0$. Conversely, $\theta^* = \underline{\theta}$ is feasible only if $b^* = 0$. If $\theta^* = \bar{\theta}$, then $\tilde{K}(\theta^*, b) = K(\bar{\theta})$ independently of b, since there is never any regulation in equilibrium, and it is then optimal to set $b^* = 0$. Assume finally that $\theta^* \in (\underline{\theta}, \bar{\theta})$. By the first-order and the complementary slackness condition for b^*:

$$b^*[\int_{\underline{\theta}}^{\theta^*} (V_k(k^*, \theta) + \Pi_k(k^*)) dF(\theta) - R_k(k^*) - \lambda^* \Theta_k(k^*)] \tilde{K}_b(\theta^*, b^*) = 0.$$

Because $\tilde{K}_b(\tilde{\theta}, b) > 0$ if $\tilde{\theta} \in (\underline{\theta}, \bar{\theta})$, it follows that $b^* > 0$ only if

$$\int_{\underline{\theta}}^{\theta^*} (V_k(k^*, \theta) + \Pi_k(k^*)) dF(\theta) - R_k(k^*) - \lambda^* \Theta_k(k^*) = 0.$$

Then the marginal net benefit of increasing $\tilde{\theta}$ above θ^*,

$$\begin{aligned}&[V(k^*,\theta^*)+\Pi(k^*)]f(\theta^*)+\lambda^*\\&+[\int_{\underline{\theta}}^{\theta^*}(V_k(k^*,\theta)+\Pi_k(k^*))dF(\theta)-R_k(k^*)-\lambda^*\Theta_k(k^*)]\tilde{K}_\theta(\theta^*,b^*),\end{aligned}$$

is strictly positive for all $\theta^* < \Theta^J(k^*)$ because then $V(k^*,\theta^*)+\Pi(k^*) > 0$. It is strictly negative for all $\theta^* > \Theta^J(k^*)$ because then $V(k^*,\theta^*)+\Pi(k^*) < 0$ and $\lambda^* = 0$ by $\Theta^J(k^*) > \Theta(k^*)$. Hence, $b^* > 0$ only if $\theta^* = \Theta^J(k^*)$, in which case also $\lambda^* = 0$. By the strict quasi-concavity of $\Omega(k)$,

$$\int_{\underline{\theta}}^{\Theta^J(k^*)}(V_k(k^*,\theta)+\Pi_k(k^*))dF(\theta)-R_k(k^*)=0$$

has a unique solution, namely $k^* = k^J > 0$. In this case, $\theta^* = \Theta^J(k^*) = \Theta^J(k^J) = \theta^J$. Using the first-order condition for profit maximization,

$$(1-F(\theta^J))b^* = R_k(k^J) - F(\theta^J)\Pi_k(k^J).$$

We next show that the right-hand side of this expression is strictly negative, which yields $b^* < 0$ and therefore a contradiction. Stability condition (11) implies that for every solution k' satisfying $R_k(k') - F(\Theta^J(k'))\Pi_k^j(k') = 0$:

$$\begin{aligned}&\frac{d}{dk}[R_k(k)-F(\Theta^J(k))\Pi_k(k)]|_{k=k'}\\=\;&[R_{kk}(k')-F(\Theta^J(k'))\Pi_{kk}(k)][1-\Theta^J_k(k')K_\theta(\Theta^J(k'))]>0.\end{aligned}$$

Hence, k^E uniquely defines k'. Furthermore, $R_k(k) - F(\Theta^J(k))\Pi_k(k) < 0$ for all $k < k^E$. Also, $\Omega_k(k^E) = X(k^E,\theta^E) < 0$ implies $k^J < k^E$. Combining these results gives $R_k(k^J) - F(\theta^J)\Pi_k(k^J) < 0$. We conclude that $b^* = 0$ also for $\theta^* \in (\underline{\theta},\bar{\theta})$ under the assumption that $X(k^E,\theta^E) < 0$.

The optimal choice of $\tilde{\theta}$: Observe that $\tilde{K}(\tilde{\theta},b^*) = \tilde{K}(\tilde{\theta},0) = K(\tilde{\theta})$. Suppose $\tilde{\theta} \geq \theta^E$ so that $\tilde{\theta} \geq \Theta^J(K(\tilde{\theta}))$. In this case, the expected joint surplus of the two contracting parties satisfies

$$\begin{aligned}\tilde{\Omega}(\tilde{\theta},0) &= \Omega(K(\tilde{\theta})) + \int_{\Theta^J(K(\tilde{\theta}))}^{\tilde{\theta}}[V(K(\tilde{\theta}),\theta)+\Pi(K(\tilde{\theta}))]dF(\theta)\\&\leq \Omega(K(\tilde{\theta}))\\&\leq \Omega(k^E)\\&= \tilde{\Omega}(\theta^E,0),\end{aligned}$$

where the first inequality follows from $V(K(\tilde{\theta}),\theta)+\Pi(K(\tilde{\theta})) < 0$ for all $\theta > \Theta^J(K(\tilde{\theta}))$. The second inequality holds because k^E maximizes $\Omega(k)$ over $[k^E, K(\bar{\theta})]$ if $X(k^E,\theta^E) < 0$. Hence, $\theta^* \leq \theta^E$. In fact, $\theta^* < \theta^E$ because $X(k^E,\theta^E) < 0$ implies $\tilde{\Omega}_\theta(\theta^E,0) = \Omega_k(k^E)K_\theta(\theta^E) < 0$. For $\tilde{K}(\tilde{\theta},b^*) = K(\tilde{\theta})$,

$\tilde{\theta} \geq \Theta(\tilde{K}(\tilde{\theta}, b^*))$ is satisfied if and only if $\tilde{\theta} \geq \theta^0$ by stability condition (6). Hence, θ^* solves

$$\max_{\tilde{\theta} \in [\theta^0, \theta^E]} \{\int_{\underline{\theta}}^{\tilde{\theta}} [V(K(\tilde{\theta}), \theta) + \Pi(K(\tilde{\theta}))] dF(\theta) - R(K(\tilde{\theta}))\} = \max_{\tilde{\theta} \in [\theta^0, \theta^E]} \tilde{W}(\tilde{\theta})$$

for $X(k^E, \theta^E) < 0$, which has solution $\theta^* = \theta^N$. This is exactly the same solution as under compensation rule (10). Because compensation payments are lump-sum, the expected joint surplus of an agreement with investment protection level θ^N sustained by (10) is the same as one sustained by $\hat{T}(k, \theta^N, 0)$, namely $\tilde{W}(\theta^N) = \tilde{\Omega}(\theta^N, 0) = \tilde{\Omega}(\theta^*, b^*) \geq \tilde{\Omega}(\hat{\theta}, \bar{b})$. We conclude that (10) maximizes the joint expected surplus of the two contracting parties among all compensation functions that are non-decreasing in k, if $X(k^E, \theta^E) < 0$.■

A.5 Proof of Proposition 5

Part *(i)*: Let $\theta' \in [\theta^0(\alpha), \theta^E(\alpha)]$. Assume that the firm perceives the regulatory threshold $\hat{\theta} = \theta'$ and therefore invests $K(\theta', \theta', \alpha)$. The host country net benefit of allowing production then equals

$$\begin{aligned} & V(K(\theta', \theta', \alpha), \theta) + (1 + \alpha)\Pi(K(\theta', \theta', \alpha)) \\ = \; & V(K(\theta', \theta', \alpha), \theta) - V(K(\theta', \theta', \alpha), \Theta^G(K(\theta', \theta', \alpha), \alpha)) \end{aligned}$$

for $\theta \leq \theta'$. By an extension of stability condition (11), $\theta' \leq \Theta^J(K(\theta', \theta', \alpha)) \leq \Theta^G(K(\theta', \theta', \alpha), \alpha)$, so the host country allows production if $\theta \leq \theta'$. Instead, the net benefit of allowing production equals

$$\begin{aligned} & V(K(\theta', \theta', \alpha), \theta) + \alpha\Pi(K(\theta', \theta', \alpha)) \\ = \; & V(K(\theta', \theta', \alpha), \theta) - V(K(\theta', \theta', \alpha), \Theta(K(\theta', \theta', \alpha), \alpha)) \end{aligned}$$

for $\theta > \theta'$. By an extension of stability condition (6), $\theta' \geq \Theta(K(\theta', \theta', \alpha), \alpha)$ for all $\theta' \geq \theta^0(\alpha)$. Hence, it is optimal for the host country to regulate for $\theta > \theta'$. The regulatory threshold $\hat{\theta} = \theta'$ as perceived by the firm therefore is consistent. It also implies that there will be regulation for all $\theta > \Theta^J(K(\theta', \theta', \alpha))$ and therefore no joint regulatory chill.

Part *(ii)*: Let $\theta' > \theta^E(\alpha)$, but $\theta' \leq \Theta^G(K(\theta', \theta', \alpha), \alpha)$. Assume that the firm perceives a regulatory threshold $\hat{\theta} = \theta'$ and therefore invests $K(\theta', \theta', \alpha)$. By arguments analogous to the ones above, the net benefit of allowing production is non-negative if and only if $\theta \leq \theta'$. Hence, the perceived regulatory threshold is consistent, and there will be joint regulatory chill for $\theta \in (\Theta^J(K(\theta', \theta', \alpha)), \theta']$. Next, let $\theta' > \Theta^G(K(\theta', \theta', \alpha), \alpha)$. In this case, $\hat{\theta} = \theta'$ would be inconsistent because the regulatory threshold implemented by the host country would be $\hat{\theta} = \Theta^G(K(\theta', \theta', \alpha), \alpha) < \theta'$. Let $\hat{\theta} < \theta'$ and such that $\theta' > \Theta^G(K(\theta', \hat{\theta}, \alpha))$. The net benefit of allowing production is non-negative if and only if $\theta \leq \Theta^G(K(\theta', \hat{\theta}, \alpha), \alpha)$. Hence, $\hat{\theta} = \Theta^G(K(\theta', \hat{\theta}, \alpha), \alpha) < \theta'$ is the threshold for regulation. In this case, there will be joint regulatory chill for all $\theta \in (\Theta^J(K(\theta', \hat{\theta}, \alpha)), \Theta^G(K(\theta', \hat{\theta}, \alpha), \alpha)]$.■

A.6 Proof of the statement in footnote 27

Let $\hat{\theta}^B$ be the firm's consistent belief about investment protection and $\hat{k}^B \equiv K(\hat{\theta}^B)$ its profit maximizing investment subsequent to an announcement of $\theta' \leq \theta^0$ as the level of investment protection in the agreement. The firm earns its full operating profit if $\theta \leq \max\{\theta'; \Theta(\hat{k}^B)\}$ and obtain zero profit otherwise. Hence, the firm's beliefs about investment protection is consistent with host country regulation only if $\hat{\theta}^B \in \{\theta'; \theta^0\}$ because $\hat{\theta}^B = \Theta(\hat{k}^B)$ if and only if $\hat{\theta}^B = \theta^0$ by assumption (6). Assume that $\theta' < \theta^0$ and suppose $\hat{\theta}^B = \theta'$. In this case, the host country optimally permits production if and only if $\theta \leq \Theta(K(\theta')) > \theta' = \hat{\theta}^B$, which is inconsistent. Hence, the only candidate for consistent beliefs is $\hat{\theta}^B = \theta^0$ for $\theta' \leq \theta^0$. The optimal investment then equals $k^0 = K(\theta^0)$, and the threshold for regulation occurs at $\Theta(k^0) = \theta^0$, which verifies consistency in this final case. Hence, all agreements with $\theta' \leq \theta^0$ would implement the Nash equilibrium (k^0, θ^0) and therefore fail to generate any net value. ■

A.7 Proof of Lemma 2

Consider first the properties of θ^N_{NTOnly}. Observe that

$$\tilde{V}(\theta') + \tilde{W}(\theta') = 2\tilde{V}(\theta') + \tilde{\Pi}(\theta')$$

implies a welfare difference

$$\tilde{V}(\theta^U) + \tilde{W}(\theta^U) - \tilde{V}(\theta') - \tilde{W}(\theta') = 2[\tilde{V}(\theta^U) - \tilde{V}(\theta')] + \tilde{\Pi}(\theta^U) - \tilde{\Pi}(\theta'),$$

which is strictly positive for all $\theta' < \theta^U$. Hence, $\theta^N_{NTOnly} \geq \theta^U$. Alternatively,

$$\tilde{V}(\theta') + \tilde{W}(\theta') = 2\tilde{W}(\theta') - \tilde{\Pi}(\theta'),$$

which implies a welfare difference

$$\tilde{V}(\theta^N) + \tilde{W}(\theta^N) - \tilde{V}(\theta') - \tilde{W}(\theta') = 2[\tilde{W}(\theta^W) - \tilde{W}(\theta')] + \tilde{\Pi}(\theta') - \tilde{\Pi}(\theta^N),$$

which is strictly positive for all $\theta' > \theta^N$. Hence, $\theta^N_{NTOnly} \leq \theta^W$. The inequalities are strict if $\theta^N_{NTOnly} \in (\theta^0, \bar{\theta})$. It is obviously the case that $\theta^N_{NTOnly} > \theta^U$ if $\theta^U = \theta^0$. But $\theta^N_{NTOnly} > \theta^U$ also if $\theta^U > \theta^0$ because then

$$\tilde{V}_\theta(\theta^U) + \tilde{W}_\theta(\theta^U) = \tilde{\Pi}_\theta(\theta^U) > 0.$$

Similarly, $\theta^N_{NTOnly} < \theta^W$ if $\theta^W = \bar{\theta}$, but $\theta^N_{NTOnly} < \theta^W$ also if $\theta^W < \bar{\theta}$ because then

$$\tilde{V}_\theta(\theta^W) + \tilde{W}_\theta(\theta^W) = -\tilde{\Pi}_\theta(\theta^W) < 0.\blacksquare$$

A.8 Proof of Proposition 9

Assume that $X(k^E, \theta^E) < 0$, so that $k^J < k^E = K(\theta^E)$; see Appendix A.4. Assume also that the Nash equilibrium yields underinvestment: $K(\theta^0) = k^0 < k^J$. There exists then a $\theta' \in (\theta^0, \theta^E)$ such that $K(\theta') = k^J$. By stability conditions (6) and (11), $\theta' \in (\theta^0, \theta^E)$ implies $\theta' > \Theta(K(\theta')) = \Theta(k^J)$ and $\theta' < \Theta^J(K(\theta')) = \Theta^J(k^J) = \theta^J$. Consider the following compensation rule:

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

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 $\theta \leq \theta' < \theta^J$. Assume that firms have invested k^J. 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 $\theta' < \theta^J$, which is the critical value above which the host country is willing to pay full compensation in order to terminate production for the investment k^J. For $\theta' < \theta \leq \theta^J$, the host country will not regulate since it then has to pay full compensation. But since it can expropriate directly without compensation, it will do so instead. For $\theta > \theta^J$ it will regulate, and not pay any compensation. Hence, given the investment k^J production will be maintained for $\theta \leq \theta^J$, which is globally efficient.

Investors 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^J fulfills the first–order condition (2).$\blacksquare$

A.9 Proof of Proposition 10

We can without loss of generality restrict attention to threshold functions that satisfy $\tilde{\Theta}(k) \geq \Theta(k)$ for all $k \geq 0$ because the home country would allow production for all $\theta \in (\tilde{\Theta}(k), \Theta(k)]$ for any non-negative compensation scheme. The expected profit of the firm then equals $F(\tilde{\Theta}(k))\Pi(k) - R(k)$ under the compensation scheme described in the Proposition. Let $\pi^J = F(\tilde{\Theta}(k^J))\Pi(k^J) - R(k^J)$ be the firm's expected profit if it invests k^J. Consider first necessity. If $\Pi(k^J) - R(k^J) < \pi^{SPE}$, then it is strictly profitable for the firm to deviate from k^J to k^{SPE}:

$$\begin{aligned} F(\tilde{\Theta}(k^{SPE}))\Pi(k^{SPE}) - R(k^{SPE}) &= [F(\tilde{\Theta}(k^{SPE})) - F(\theta^{SPE})]\Pi(k^{SPE}) + \pi^{SPE} \\ &\geq \pi^{SPE} > \Pi(k^J) - R(k^J) = (1 - F(\tilde{\Theta}(k^J)))\Pi(k^J) + \pi^J \geq \pi^J. \end{aligned}$$

To show sufficiency, define the threshold function

$$\begin{aligned}\tilde{\Theta}(k) &= \max\{y - \frac{[F(y)-F(\theta^J)]\Pi_k(k^J) - X(k^J,\theta^J)}{f(y)\Pi(k^J)} h(k); \Theta(k)\},\\ y &= \max\{\theta^J; F^{-1}\left(\frac{\pi^{SPE}+R(k^J)}{\Pi(k^J)}\right)\},\ h(k^J)=0,\ h'(k^J)=1.\end{aligned}$$

$y \leq \bar{\theta}$ by the assumption that $\Pi(k^J) - R(k^J) \geq \pi^{SPE}$. To find the profit-maximizing investment, let

$$\kappa = \{k \geq 0 | y - \frac{[F(y)-F(\theta^J)]\Pi_k(k^J) - X(k^J,\theta^J)}{f(y)\Pi(k^J)} h(k) \geq \Theta(k)\}$$

and denote its complement by $\kappa^c = R_+ \backslash \kappa$. By construction, $k^J \in \kappa$:

$$y - \frac{[F(y)-F(\theta^J)]\Pi_k(k^J) - X(k^J,\theta^J)}{f(y)\Pi(k^J)} h(k^J) = y \geq \theta^J = \Theta^J(k^J) > \Theta(k^J).$$

It cannot be strictly profitable for the firm to deviate to $k \in \kappa^c$ because then

$$\begin{aligned}F(\tilde{\Theta}(k))\Pi(k) - R(k) &= F(\Theta(k))\Pi(k) - R(k) \leq \pi^{SPE}\\ &= F\left[F^{-1}\left(\frac{\pi^{SPE}+R(k^J)}{\Pi(k^J)}\right)\right]\Pi(k^J) - R(k^J) \leq F(y)\Pi(k^J) - R(k^J) = \pi^J.\end{aligned}$$

We next verify that k^J is an optimal investment choice within κ. Every profit-maximizing investment level $k > 0$ in this domain must satisfy

$$\{X(k^J,\theta^J) - [F(y)-F(\theta^J)]\Pi_k(k^J)\}\frac{f(\tilde{\Theta}(k))}{f(y)}\frac{\Pi(k)}{\Pi(k^J)} h'(k) + F(\tilde{\Theta}(k))\Pi_k(k) - R_k(k) = 0.$$

Obviously, $k = k^J$ solves this first-order condition. The corresponding second-derivative of the profit function is

$$\left[\frac{h''(k)}{h'(k)} + \frac{f'(\tilde{\Theta}(k))}{f(\tilde{\Theta}(k))}\tilde{\Theta}'(k) + 2\frac{\Pi_k(k)}{\Pi(k)}\right]\Pi(k) f(\tilde{\Theta}(k))\tilde{\Theta}'(k) + F(\tilde{\Theta}(k))\Pi_{kk}(k) - R_{kk}(k).$$

We can choose $h''(k)$ such that $F(\tilde{\Theta}(k))\Pi(k) - R(k)$ is strictly concave across the entire domain κ, which establishes k^J as the profit-maximizing investment under the specific threshold function considered in this proof. The equilibrium level of investment protection equals $\tilde{\Theta}(k^J) = y$. If $y = \theta^J$, then it is optimal for the home country to allow production for all $\theta \leq \theta^J$ and to regulate otherwise. If $y > \theta^J$, then it is still optimal for the home country to allow production for $\theta \leq \theta^J$ and to regulate otherwise. In this final case, the country pays out compensation for all $\theta \in (\theta^J, y]$. Either way, the equilibrium threshold for regulation is θ^J.■

A.10 Proof of Proposition 11

Building on the generalized model introduced in Appendix A.3, we state and prove a slightly more general result than Proposition 11:

Proposition A.3 *Assume that the industry in country i consists of a single regulation-strategic foreign firm. For any investment agreement T^i that satisfies features (a)-(c), and for which compensation at most equals foregone operating profit, there exists an alternative agreement that satisfies features (a)-(c), and that for each country i:*
(i) features the compensation function

$$\tilde{T}^i(k_i, \theta_i) \equiv \begin{cases} \Pi^i(k_i) & \text{if } \theta_i \le \tilde{\Theta}^i(k_i, \gamma_i) \\ 0 & \text{if } \theta_i > \tilde{\Theta}^i(k_i, \gamma_i); \end{cases} \tag{A.14}$$

(ii) implements a threshold function for regulation $\hat{\Theta}^i(k_i, \gamma_i) = \min\{\tilde{\Theta}^i(k_i, \gamma_i); \Theta^{iJ}(k_i)\} \ge \Theta^i(k_i, \gamma_i)$;
(iii) yields weakly higher expected welfare in both countries and foreign industry profits than the initial agreement.

Proof: For any arbitrary $T^i(k_i, \theta_i) \in [0, \Pi^j(k_i)]$, define the threshold $\tilde{\Theta}^i(k_i, \gamma_i)$ by

$$F^i(\tilde{\Theta}^i(k_i, \gamma_i)) \equiv \int_{M^i(k_i)} dF^i(\theta_i) + \int_{M^{ir}(k_i)} \frac{T^i(k_i, \theta_i)}{\Pi^j(k_i)} dF^i(\theta_i) \le 1$$

and the compensation mechanism by (A.14).

We know from the proof of Proposition A.1 that $\int_{M^i(k_i)} dF^i(\theta_i) \ge F^i(\Theta^i(k_i, \gamma_i))$. Hence, $\tilde{\Theta}^i(k_i, \gamma_i) \ge \Theta^i(k_i, \gamma_i)$. And because also $\Theta^{iJ}(k_i) \ge \Theta^i(k_i, \gamma_i)$, it follows that $\hat{\Theta}^i(k_i, \gamma_i) \ge \Theta^i(k_i, \gamma_i)$. If $\tilde{\Theta}^i(k_i, \gamma_i) \le \Theta^{iJ}(k_i)$, then it is optimal to maintain production for all $\theta_i \le \tilde{\Theta}^i(k_i, \gamma_i)$ and to regulate for all $\theta_i > \tilde{\Theta}^i(k_i, \gamma_i)$. If $\tilde{\Theta}^i(k_i, \gamma_i) > \Theta^{iJ}(k_i)$, then it is optimal to maintain production for all $\theta_i \le \Theta^{iJ}(k_i)$ and to regulate for all $\theta_i > \Theta^{iJ}(k_i)$. Hence, country i regulates if and only if $\theta_i > \hat{\Theta}^i(k_i, \gamma_i)$.

Investments are the same under the two agreements. The monopoly receives its operating profit $\Pi^i(k_i)$ for all $\theta_i \le \tilde{\Theta}^i(k_i, \gamma_i)$ independently of whether it is regulated or not. It is regulated, but receives no compensation for all $\theta_i > \tilde{\Theta}^i(k_i, \gamma_i)$. The monopoly's expected operating profit in country i thus equals

$$\Pi^i(k_i) F^i(\tilde{\Theta}^i(k_i, \gamma_i)) = \Pi^j(k_i) \int_{M^i(k_i)} dF^i(\theta_i) + \int_{M^{ir}(k_i)} T^i(k_i, \theta_i) dF^i(\theta_i),$$

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

The proofs that welfare in country j remains the same, that all firms are equally well off as before and that welfare in country i is weakly higher under compensation rule $\tilde{T}^i$ than T^i are identical to those in Appendix 3 and therefore not repeated here.■

A.11 A compensation scheme based on relative performance

Assume that the industry in the host country consists of $H \geq 2$ symmetric foreign firms—the results hold also for some degree of firm heterogeneity. We index firms by $h \neq \hat{h} = 1, ..., H$. Let k_h be the investment of firm h and $\mathbf{k} = (k_h, \mathbf{k}_{-h})$ the investment profile of all firms, where $\mathbf{k}_{-h} = (k_1, ..., k_{h-1}, k_{h+1}, ...k_H)$ represents the investment profile of all firms other than h. We let the operating profit of firm h be $\Pi^h(\mathbf{k}) \equiv \hat{\Pi}(P(\mathbf{k}), k_h)$, where $p = P(\mathbf{k})$ represents market effects that each firms treats as exogenously given. These "price" effects are pure redistribution between the industry and the host country and therefore have no joint welfare effects. Let $\hat{\Pi}_k(P(\mathbf{k}), k_h) > 0$ be the marginal perceived effect on firm h's operating profit of increasing investment k_h. 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. Furthermore, we let

$$V_{k_k}(\mathbf{k}, \theta) + \sum_{\hat{h}=1}^{H} \Pi_{k_h}^{\hat{h}}(\mathbf{k}) = \Psi_{k_h}(\mathbf{k},\theta) + \hat{\Pi}_k(P(\mathbf{k}), k_h)$$

be the marginal joint effect of the investment on the host country and the industry of a marginal increase in k_h given the value θ of the regulatory shock. The function $\Psi(\mathbf{k},\theta)$ represents the host country externalities of the FDI; we assume that these externalities do not impact the operating profits of the firms.

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

$$V(\mathbf{k}, \Theta^J) + \sum_{h=1}^{H} \Pi^h(\mathbf{k}) \equiv 0.$$

The jointly welfare maximizing investment profile features the same investment k^J by all firms, and is the solution to

$$F(\theta^J)\hat{\Pi}_k(p^J, k^J) - R_k(k^J) + \int_{\underline{\theta}}^{\theta^J} \Psi_{k_h}(k^J\mathbf{1},\theta) dF(\theta) = 0,$$

where $\mathbf{1}$ is an H-dimensional vector of 1, the efficient threshold for regulation is $\theta^J = \Theta^J(k^J\mathbf{1})$, and $p^J = P(k^J\mathbf{1})$.

Let

$$\Delta\tilde{\Psi}^h(\mathbf{k}) \equiv \int_{\underline{\theta}}^{\Theta^J(\mathbf{k})} (\Psi(\mathbf{k},\theta) - \Psi(0, \mathbf{k}_{-h}, \theta)) dF(\theta)$$

be the expected externalities associated with firm h's investment if regulation is ex post efficient. Assume that $\Psi_{k_h k_h} \leq 0$ for all h and that each firm h treats all other firms' externalities as exogenous to the own investment k_h.[35]

[35] Independence is a behavioral assumption here, but could be affected by technology. If the externality is additive,

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

$$T^h(\mathbf{k}) = \begin{cases} \frac{1}{|\mathcal{H}(\mathbf{k})|-1} \sum_{\hat{h}\in\mathcal{H}(\mathbf{k})\setminus h}[\Pi^h(\mathbf{k}) + \frac{\Delta\tilde{\Psi}^h(\mathbf{k})-\Delta\tilde{\Psi}^{\hat{h}}(\mathbf{k})}{1-F(\Theta^J(\mathbf{k}))}] & \forall h \in \mathcal{H}(\mathbf{k}) \\ 0 & \forall h \notin \mathcal{H}(\mathbf{k}) \end{cases} \tag{A.15}$$

yields non-negative compensation for all firms in the industry. Here, the compensation depends not only on operating profit, but also on the externalities. For instance, the firm receives a relatively large compensation if the externalities of its investment are positive compared to that of the other firms in the industry.[36]

The total payment does not involve third parties nor does it ever imply overcompensating the industry by the host country for any possible investment profile $\mathbf{k}$ or realization of the shock θ:

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

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

Holding the threshold fixed at θ^J and the "price" at p^J, and assuming $\mathbf{k}_{-h} = \mathbf{k}_{-h}^J$, the expected profit of firm h is

$$F(\theta^J)\hat{\Pi}(p^J, k_h) + (1 - F(\theta^J))T^h(k_h, \mathbf{k}_{-h}^J) - R(k_h)$$

The perceived marginal effect

$$F(\theta^J)\hat{\Pi}_k(p^J, k_h) - R_k(k_h) + \int_{\underline{\theta}}^{\theta^J} \Psi_{k_h}(k_h, \mathbf{k}_{-h}^J, \theta) dF(\theta)$$

on the expected profit of increasing investment k_h is exactly the same as the marginal expected joint welfare effect. By the construction of the mechanism (A.15), the host country and the firms all internalize the full economic effects of their actions.

Proposition A.4 *Assume that there are $H \geq 2$ identical foreign firms in the industry and that each firms treats all market effects, regulation and the externalities of the other firms' investment, as exogenous to the own investment. Assume also that the operating profits at the efficient outcome are sufficiently large that $\Pi(k^J\mathbf{1}) - R(k^J) \geq \max_{k\geq 0}\{F(\theta^J)\hat{\Pi}(p^J, k) - R(k)\}$. In this case, the fully efficient outcome $(k^J\mathbf{1}, \theta^J)$ can be implemented as a Nash equilibrium by an investment agreement that stipulates compensation according to (A.15).*

$\Psi(\mathbf{k},\theta) \equiv \sum_{h=1}^{I} \hat{\Psi}^h(k_h,\theta)$, then $\Delta\tilde{\Psi}^{\hat{h}}(\mathbf{k}) = \int_{\underline{\theta}}^{\hat{\theta}}[\hat{\Psi}^{\hat{h}}(k_{\hat{h}},\theta) - \hat{\Psi}^{\hat{h}}(0,\theta)]dF(\theta)$, which is independent of k_h for all $\hat{h} \neq h$ if firm h also treats the probability $\hat{\theta}$ of regulation as exogenous to its own investment k_h.

[36] The compensation rule (A.15) is defined only for $|\mathcal{H}(\mathbf{k})| \geq 2$. For completeness, assume that the firm with the maximal $\Pi^h(\mathbf{k}) + \frac{\Delta\tilde{\Psi}^h(\mathbf{k})}{1-F(\Theta^G(\mathbf{k}))}$ is compensated by $\Pi^h(\mathbf{k})$ and that the rest of the firms receive nothing in compensation if $|\mathcal{H}(\mathbf{k})| = 1$.

Proof: The host country must pay the full industry profit in compensation if $\mathbf{k} = k^J\mathbf{1}$ and will therefore allow production if and only if $\theta \leq \theta^J$. Assume that $\mathbf{k}_{-h} = \mathbf{k}^J_{-h}$ and consider the choice of k_h under the assumption that firm h expects to be regulated with probability θ^J. By strict concavity of the profit function, $k_h = k^J$ is the profit maximizing investment in the domain $k_h \leq \kappa$, where $\kappa > k^J$ is the upper bound to firm h's investment that yields a strictly positive compensation under regulation. The expected equilibrium profit $\Pi(k^J\mathbf{1}) - R(k^J)$ by assumption is larger than the maximum profit, $\max_{k\geq 0}\{F(\theta^J)\hat{\Pi}(p^J, k) - R(k)\}$, the firm could obtain if it received no compensation. This is also a necessary condition for implementation of the fully efficient outcome under asymmetric information. Hence, k^J is firm h's profit maximizing investment for all $k_h \geq 0$. By continuity, the proposition holds also for some degree firm heterogeneity.■

Remarks Implementation of the efficient outcome is independent of any information concerning the extent to which the host country internalizes operating profit. To see this, assume that the host country attaches a weight $\gamma_h \in [0, 1]$ to the operating profit of firm h. In this case, the net benefit of allowing production equals $V(k^J\mathbf{1}, \theta) + \sum_{h=1}^{H}[\gamma_h\Pi^h(k^J\mathbf{1}) + (1 - \gamma_h)T^h(k^J\mathbf{1})]$ at the efficient investment profile $k^J\mathbf{1}$. This is equal to $V(k^J\mathbf{1}, \theta) + \sum_{h=1}^{H}\Pi^h(k^J\mathbf{1})$ under (A.15) and therefore independent of all γ_h because $T^h(k^J\mathbf{1}) = \Pi^h(k^J\mathbf{1})$ for all h.

Proposition A.4 also holds for some degree of firm heterogeneity. The important part is that firms are sufficiently similar that $|\mathcal{H}(\mathbf{k}^J)| = H$, so that the ex post incentive to regulate is efficient at $\mathbf{k}^J$. Under certain conditions, the mechanism is still efficient with larger firm differences. This happens if the industry can be partitioned into multiple comparison groups with two or more similar firms in each group, such that all of them receive positive compensation in a neighborhood around $\mathbf{k}^J$.

A.12 Proof of Proposition 12

Assume that compensation can only be paid if the firm is regulated and that it cannot exceed $\Pi(k)$. 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 scheme $\hat{T}(k,\theta)$.

The equilibrium rent of the host country is

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

By standard arguments, the compensation scheme is incentive compatible only if $\mathcal{V}_\theta(k,\theta) = \xi(\theta)V_\theta(k,\theta)$, and $\xi(\theta)$ is non-increasing in θ. The expected rent is

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

The incentive compatible compensation is therefore given by

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

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 $\mathcal{V}(k,\underline{\theta}) = V(k,\underline{\theta})$. 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

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

A threshold $\hat{\theta} < \Theta(k)$ cannot be implemented since this would imply negative compensation:

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

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

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

which we have ruled out by assumption.

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

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

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 $-V(k,\bar{\Theta}(k))$ whenever it disallows production. In this case, the net benefit $V(k,\theta) - V(k,\bar{\Theta}(k))$ of allowing production would be non-negative if and only if $\theta \leq \bar{\Theta}(k)$.

An alternative approach, which would highlight the role of asymmetric information, would be to decentralize the decision to regulate to the country, but require it to report θ and pay compensation

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

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

If instead $\hat{\theta} \in (\Theta(k), \Theta^J(k)]$, the host country would report $\widetilde{\theta} > \hat{\theta}$ subsequent to regulation because doing so would minimize the compensation payment: $-V(k,\hat{\theta}) \leq \Pi(k)$. In this case, the net benefit $V(k,\theta) - V(k,\hat{\theta})$ 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 $-V(k,\hat{\theta}) > 0$.

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