European University Institute

ROBERT SCHUMAN CENTRE FOR ADVANCED STUDIES

EUI WORKING PAPERS

RSCAS 2017/19
Robert Schuman Centre for Advanced Studies
Global Governance Programme-261

Economics and Politics of International Investment Agreements

Henrik Horn and Thomas Tangerås

European University Institute
Robert Schuman Centre for Advanced Studies
Global Governance Programme

Economics and Politics of International Investment Agreements

Henrik Horn and Thomas Tangerås

EUI Working Paper **RSCAS** 2017/19

ISSN 1028-3625

Printed in Italy, March 2017
European University Institute
Badia Fiesolana
I – 50014 San Domenico di Fiesole (FI)
Italy
www.eui.eu/RSCAS/Publications/
www.eui.eu
cadmus.eui.eu

Robert Schuman Centre for Advanced Studies

The Robert Schuman Centre for Advanced Studies (RSCAS), created in 1992 and directed by Professor Brigid Laffan, aims to develop inter-disciplinary and comparative research and to promote work on the major issues facing the process of integration and European society.

The Centre is home to a large post-doctoral programme and hosts major research programmes and projects, and a range of working groups and *ad hoc* initiatives. The research agenda is organised around a set of core themes and is continuously evolving, reflecting the changing agenda of European integration and the expanding membership of the European Union.

Details of the research of the Centre can be found on:
http://www.eui.eu/RSCAS/Research/

Research publications take the form of Working Papers, Policy Papers, Policy Briefs, Distinguished Lectures, Research Project Reports and Books.

Most of these are also available on the RSCAS website:
http://www.eui.eu/RSCAS/Publications/

The EUI and the RSCAS are not responsible for the opinion expressed by the author(s).

The Global Governance Programme at the EUI

The Global Governance Programme is one of the flagship programmes of the Robert Schuman Centre for Advanced Studies at the European University Institute (EUI). It aims to: build a community of outstanding professors and scholars, produce high quality research and, engage with the world of practice through policy dialogue. At the Global Governance Programme, established and early career scholars research, write on and discuss, within and beyond academia, issues of global governance, focussing on four broad and interdisciplinary areas: European, Transnational and Global Governance; Global Economics; Europe in the World; and Cultural Pluralism.

The Programme also aims to contribute to the fostering of present and future generations of policy and decision makers through its unique executive training programme, the Academy of Global Governance, where theory and “real world” experience meet. At the Academy, executives, policy makers, diplomats, officials, private sector professionals and academics, have the opportunity to meet, share views and debate with leading academics, top-level officials, heads of international organisations and senior executives, on topical issues relating to governance.

For more information: http://globalgovernanceprogramme.eui.eu

Abstract

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

Keywords

Foreign direct investment; expropriation; international investment agreements; regulatory chill.

JEL Classification: F21; F23; F53; K33

ECONOMICS AND POLITICS 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: January 27, 2017

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

[2]Corresponding author. Address: Research Institute of Industrial Economics (IFN); Box 55665; SE-102 15 Stockholm; Sweden. Phone: +46 70 742 88 39; Fax: +46 8 665 45 99; Email address: henrik.horn@ifn.se.

[3]Email address: thomas.tangeras@ifn.se.

1 Introduction

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

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

Investment agreements have until recently been formed without much political opposition, but some agreements have recently come under intense fire. The debate has concerned in particular the role of investment protection in "mega-regional" preferential trade agreements—the Trans-Pacific Partnership (TPP), the Canada-EU Comprehensive Economic and Trade Agreement (CETA), and the EU-US Transatlantic Trade and Investment Partnership (TTIP). Central to this critique is the fact that these agreements cover not only direct expropriation, but also cases where host countries take regulatory measures that significantly reduce the value of the investment to its owners without taking over ownership of the assets—so called *indirect* (or *regulatory*) expropriation. Critics argue that these (and other) provision are so generally formulated that almost any regulatory policy with adverse consequences for foreign investors could be interpreted to require compensation. This is seen to be particularly troublesome given the possibility for investors to use the very potent ISDS mechanisms to enforce the substantive obligations in the agreements. It is argued that signatory states will refrain from pursuing legitimate public policy goals to avoid litigation, that is, the agreements will cause "regulatory chill."[2]

This skeptical view has been fuelled by a number of actual litigations that have made headlines: TransCanada Corporation recently declared its intention to sue the US under NAFTA for US$ 15 billion as compensation for the decision by the Obama Administration to disallow the construction of the Keystone XL pipe line; Phillip Morris has litigated against several countries over the tobacco plain packaging legislation; Spain is facing a large number of litigations for the withdrawal of renewable energy support schemes during the financial crisis of 2008, and similar cases have been brought against Italy and the Czech Republic; Germany is being sued by the energy company

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

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

Vattenfall for losses caused by the country's decision to shut down nuclear power in the wake of the Fukushima disaster.

The policy debate raises a number of questions: How should investment agreements be designed? Do appropriately designed agreements cause regulatory chill? Do they resolve the perceived overregulation and underinvestment problems? Who benefits and who loses from the formation of the agreements? The economic literature sheds very little light on these issues. The purpose of this paper is to contribute to filling this lacuna by analyzing the optimal design and implications of agreements that compensate investors for regulatory expropriations.

An investment agreement need to address the interaction between two distortions in the case of regulatory expropriations: On the one hand, the host country disregard of foreign investor interests when deciding on regulation—this is what causes a tendency toward overregulation, and thus a potential benefit from protection—and on the other hand, the potential adverse effects from investment, which motivates the existence of a regulatory regime. The interaction between these distortions can cause overregulation and underinvestment, and thus creates a scope for an investment agreement. To capture this interaction, we consider an investment agreement between two countries in a generalized version of the canonical regulatory takings model of investment protection.[3] One central feature of our approach is that we focus on voluntary, Pareto efficient investment agreements. Another is that we for the most part constrain the agreements under study to share certain basic features with actual agreements. This is in our view a natural starting point for the analysis of investment agreements, but as explained below, the small related literature on investment agreements has mainly followed different approaches in this regard.

At the outset of the interaction, competitive firms make irreversible foreign direct investment in production facilities. The investments create benefits to the host country, but might also create negative externalities of unknown magnitude at the time of the investment. These externalities can render production ex post undesirable from a domestic, and possibly also from a joint welfare perspective. The shock could capture a broad range of exogenous events, for instance the arrival of information concerning adverse environmental or heath consequences of production, but we will simply denote the realizations as "regulatory shocks." Upon observing its country-specific shock, each host country decides whether to permit or to disallow production. Production and consumption occurs if there is no regulation. The private investment is effectively lost if the country instead regulates the industry, although there is no formal take-over of the ownership of the assets. The host country disregards the adverse effects for investors. The outcome is likely to be inefficient absent investment protection, and it might feature too little investment and excessive regulation relative to the first best. Hence, there appears to be scope for an investment agreement.

We represent investment agreement as a set of negotiated rules specifying payments to foreign investors as a function of regulatory decisions, and possibly other factors. Importantly, we require

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

compensation schemes to be congruent with core features of actual agreements. For instance, firms are only eligible for compensation in case of regulation. Compensation must be non-negative and based on, but cannot exceed firms' foregone operating profits. But we also identify plausible circumstances under which our imposed contract restrictions do not constrain agreements in terms of efficiency.

The first issue to be addressed is the key policy question of whether investment agreements cause regulatory chill. We demonstrate that Pareto optimal investment agreements never yield underregulation from a joint welfare perspective—there will be no "global regulatory chill." When compensation is limited to at most each firm's operating profit, it is always optimal for the host country to regulate when doing so is ex post efficient. Hence, agreements induce either ex post efficient regulation or inefficient overregulation. They do yield less regulation than would result without any agreement, but such "domestic regulatory chill" is simply the price the host country must pay to promote foreign investment. These results hold for a large variety of settings, for instance, they are independent of the market structure and only requires that compensation be non-negative.

A related policy issue is when investors should, and when they should not, be compensated in case of regulation. We show that a simple "carve-out" scheme by which firms receive full compensation for all regulatory shocks below, and no compensation above, a specified threshold, is sufficient to implement any Pareto optimal investment agreement, when compensation is required to be proportional to operating profit. This threshold level is referred to as the *level of investment protection*. This finding simplifies the subsequent analysis of the formation and the implications of investment agreements considerably, since the agreements can be characterized in terms of the level of investment protection they offer.

We next consider the capacity of an investment agreement to address the distortions in our setting under ideal circumstances, that is, if designed to maximize the joint welfare of the parties. We identify a non-trivial set of circumstances under which an agreement can implement the unconstrained efficient outcome, despite fulfilling our contract restrictions.

To shed light on the politics of investment agreements we then examine the distribution of the benefits and costs from an agreement. To this end we distinguish between two archetypical forms of agreements. One is a "North-South agreement" between a developed and a developing country, that serves to stimulate investment from North to South.[4] We assume that South is unable to make credible unilateral commitments to investment protection. Such an agreement, if formed, by necessity increases domestic welfare in South—that is, the welfare generated in, and accruing to, South. These benefits effectively emanate from the external legal enforcement of investment protection commitments that the agreement offers: South would have nothing to gain from an

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

agreement if it already had full commitment possibilities because then its unilaterally chosen level of investment protection would internalize all relevant effects of FDI. This mechanism corresponds closely to the "commitment approach" to trade agreements, which sees trade agreements as a tool to help governments withstand domestic protectionist pressures.

The other archetypical treaty is a "North-North agreement." Such an agreement does not improve commitment possibilities. The gains instead stem from the bilateral internalization of the externalities from regulation that the agreement allows. This mechanism is much in line with the standard view of trade agreements, which sees these agreements as means of escaping Prisoner Dilemma-like situations. The improved investment protection that a North-North agreement offers benefits investors in both countries, since the agreement entails more investment protection than is optimal from a domestic welfare point of view in both countries.

We believe these findings are informative regarding the politics of investment protection. The results suggest that symmetric agreements such as CETA and TTIP (and to some extent also the TPP) would benefit foreign investors, but reduce consumer welfare in a broad sense of the term. This might explain why the industry in general favors these agreements, while at the same time there is considerable popular resistance to their formation. Our findings also predict that there should be less opposition to North-South agreements, since the benefits to a larger extent accrue to the broader public in the host country. Again, this seems broadly compatible with what is observed, in that there appears to be much less popular discontent with investment agreements in developing countries than with the mega-regional agreements.

Yet another contentious policy issue with regard to the design of compensation mechanisms in investment agreements, is whether changes in political preferences should be treated as any other risk in investment agreements. A simple reformulation of our model allows us to examine certain aspects of this issue. We assume that there are two types of shocks that affect regulation: an exogenous regulatory shock, such as a scientific discovery, and a shock to political preferences concerning the regulatory objective, such as a change in government. Government preferences are unknown at the investment stage, but are realized simultaneously with the scientific shock. We show that an ex ante optimal investment agreement that is formed behind a "veil of ignorance," allows governments that are more sensitive to regulatory shocks to intervene for a larger range of shocks without paying compensation. Optimal agreements are thus sensitive to "democratic" concerns in this regard.

The remainder of the paper extends the analysis in a number of directions. First, investment agreements typically include provisions that prohibit less favorable treatment of foreign than of domestic investment when these are undertaken in "like circumstances." We show that host governments can benefit from including such National Treatment clauses in investment agreements as a commitment tool to enforce stricter regulation of domestic industries.

Second, investment agreements typically include stricter rules for direct expropriation compared to the rules for regulatory expropriations, in that the former often does not allow the general exceptions that apply to the latter type of expropriations. However, allowing uncompensated direct expropriation can actually enhance efficiency, since this helps separate the problem of correcting investment incentives from that of ensuring ex post optimal regulation incentives: Direct expropriation and regulation have the same consequences for the targeted firms. But once a government has seized an asset it will effectively internalize the consequences of its regulatory decisions for the return from the asset. This mechanism depends crucially on the regulatory shock being observable, however. Under asymmetric information, a host country would always claim that the realization of the shock allows it to seize assets without compensation. Full compensation for all direct expropriation is the easiest way to avoid private investment being driven completely out of the market in this case.

In a third extension we identify circumstances under which an investment agreement can implement a fully efficient outcome when the industry consists of a strategically behaving monopoly.

A fourth extension assumes that the regulatory shock is private information to the host country. A mechanism with zero compensation for large shocks is then not incentive compatible, since the host country could regulate at no domestic cost by exaggerating the severity of the regulatory shock. Incentive compatibility instead requires the country to pay a fixed compensation for all regulation. This compensation would generally depend on the cost to the host country of allowing production and not on foregone operating profit.

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

1.1 Relation to the literature

The informal economic literature on expropriation of foreign investment dates back at least to Keynes (1924), and includes e.g. Vernon's (1971) "obsolescing bargaining" theory. Eaton and Gersovitz (1983, 1984) are among the first to study expropriation in a conventional neoclassical framework. Several contributions focus on implicit mechanisms rather than international treaties for reducing investor-state hold-up problems. For instance, Dixit (1988) informally sketches an incomplete information model of the interaction between a sequence of potential investors and a host country, in a situation where the host country preferences regarding expropriation are unknown to investors (Raff, 1992, formally analyzes such a setting). Dixit (1988) shows how a host country that would benefit in the short run from expropriating might refrain from doing so in an effort to persuade investors it has other preferences. Cole and English (1991) show how the incentives for expropriation can be kept at bay by the use of trigger strategies in an infinite horizon model. Thomas and Worral (1994) and Schnitzer (1999) examine how other forms of self-enforcing agreements between investors and host governments can remedy hold-up problems. Dixit (2011) discusses a range of issues related to insecurity of property rights and FDI, and also provides extensive reviews of both the theoretical and the empirical literature.

There are two strands of mostly very recent literature that directly address the design and impact of specific obligations in investment agreements.[5] One strand considers implications of exogenously imposed investment agreements. Janeba (2016) analyzes whether national or international courts should arbitrate disputes between foreign investors and states. The paper focuses on two sources of deficiencies in the arbitration process under investment agreements. First, litigation costs can dissuade host countries from pursuing efficient policies, causing a form of regulatory chill. Second, international courts are more prone to rule in favor of investors than domestic courts. Janeba (2016) shows how the incentive for a country to form an investment agreement depends on the losses from unfavorable determinations, and the benefits for its own foreign investors from discrimination in their favor in the partner country.

Kohler and Stähler (2016) examine consequences of a particular interpretation of the "legitimate expectations" notion that sometimes has been employed by arbitration panels. It holds that past regulatory policies can create legitimate investor expectations about subsequent regulations, and thus effectively link regulatory decisions across time. The authors show in a two-period framework how such intertemporal linkages can reduce overregulation and increase aggregate welfare over time. This type of agreement is then compared with an agreement that instead imposes a National Treatment rule that equalizes the protection of foreign and domestic firms. The authors identify circumstances under which this non-discrimination rule yields higher aggregate welfare than the former mechanism.

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

Konrad (2016) considers the strategic incentives to invest in order to reduce the probability of environmental regulation. Increased investment protection benefits investors, but exacerbates an already existing overinvestment and underregulation problem. Konrad (2016) sees these results as one explanation for why firms favor investment protection and why those mechanisms are disliked by environmentalists and other interest groups.

[5] There are other, but to this paper more tangential contributions. For instance, Markusen (1998, 2001) discusses pros and cons of investment agreements from a developing country perspective. Turrini and Urban (2000, 2008) analyze the role of a multilateral investment agreement. Bergstrand and Egger (2013) depict a investment agreement as an exogenous reduction in the capital cost of FDI in a three-factor, three country general equilibrium setting. It is shown how the welfare gains of investment agreements and preferential trading agreements depend on factors such as country size and trade costs.

The welfare analyses in these papers in each case involves a comparison of an investment agreement with exogenously given characteristics to some outside option. There is no analysis of whether an alternative design of the agreement could make it welfare enhancing and acceptable to the parties if the proposed agreement is not, nor is there any discussion about the relevant alternative. For instance, it is not clear why the host country would enter into an agreement in Konrad's (2016) model in light of the fact that increased investment protection reduces host country welfare. Conversely, even if the studied agreements would increase aggregate welfare they might still not come about, since they could have adverse consequences for some of the parties. Our paper differs from the above contributions by considering the endogenous formation of an agreement that fulfills realistic contract restrictions and accounts for reasonable participation constraints of the contracting parties.

The second strand of literature examines the optimal design of investment agreements. Aisbett et al (2010a) incorporate an imperfectly unobservable regulatory shock in a standard regulatory takings model. Key contributions to the takings literature implicitly assume that the incentives to invest and to regulate are undistorted.[6] Aisbett et al (2010a) show how full efficiency can be achieved also with distorted incentives to regulate if the host country can overcompensate the industry for its losses. Aisbett et al (2010b) examine the implications of a National Treatment (NT) provision that prevents a host country from requesting up-front payments from foreign firms prior to investing. An NT rule can render broader exemptions from the compensation requirements desirable in case of regulation. Stähler (2016) derives a mechanism that can achieve the fully efficient outcome when the regulatory shock is unobservable. Efficiency is achieved by breaking the payment balance between the host country and firms.

Our paper differs from these three contributions in several regards. Most importantly, all three studies rely on compensation schemes that are not found in actual investment agreements: subsidization and overcompensation are typically not part of treaties, and the agreements do not give any scope for breaking the budget balance by payments to or from third parties. Furthermore, none of these three papers discuss distributional effects of compensation schemes, and consequently not the incentives to form the agreements.

2 Salient features of investment agreements

The investment agreements under scrutiny here should be distinguished from state-to-state tax treaties and standard commercial contracts formed between a host country and an individual investor. There is currently no multilateral investment agreement, despite the attempt by the OECD to launch such an agreement in 1998; the World Trade Organization Agreement contains certain

[6]See Blume et al (1984) and Miceli and Segerson (1994). In such instances, a compensation mechanism can only reduce welfare. Hermalin (1995) demonstrates how taxes and other sophisticated compensation mechanisms can achieve the first-best outcome in a takings model with distorted investment and regulation decisions. See Miceli and Segerson (2011) for a comprehensive survey. We discuss some of these papers in more detail in Section 5.5.

provision regarding trade-related investment, but no protection against direct or indirect expropriations. The very large number of investment agreements in force differ in coverage, and the associated case law is highly fragmented, with similar provisions interpreted very differently by different panels. Nevertheless, the more prominent negotiated agreements, such as NAFTA, the mega-regional agreements, and the EU and US "model agreements," share certain features.[7]

First, most agreements address the treatment of investments after establishment without providing pre-establishment rights.[8]

Second, the agreements almost always request non-discriminatory treatment of foreign investment (and sometimes also investors) in the sense that host country treatment of foreign investment should be "no less favorable than that it accords, in like circumstances" to its own investors (National Treatment), or to third country investors (Most-Favored Nation Treatment).[9,10]

Third, investment agreements typically specify that foreign investment should be given at least a "minimum standard of treatment." A common part of these undertakings is a commitment to provide "fair and equitable treatment." The vagueness of this concept has caused a number of contentious interpretations in case law.

Fourth, investment agreements almost invariably distinguish between direct and "indirect" expropriation, with the latter referring to "an action or series of actions by a Party [that] has an effect equivalent to direct expropriation without formal transfer of title or outright seizure." A common restriction is that expropriations are only allowed if they are for a public purpose, are non-discriminatory, are in accordance with due process of law, and if investors receive "prompt, adequate, and effective compensation." Agreements occasionally provide further specifications that provide guidance for the interpretation of "indirect expropriation". For instance, "the extent to which the government action interferes with distinct, reasonable investment-backed expectations" or "the character of the government action" determines whether a state intervention represents an expropriation. Investment agreements also increasingly include restrictions on the ambit of the indirect expropriation clauses, so called "carve-outs.". For instance:

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

and

---

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

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

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

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

> "[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."

A standard specification concerning the required magnitude for compensation in case of expropriation is that it should be equivalent to the "fair market value" of the expropriated investment. When interpreting this and related concepts, arbitration panels normally seek guidance in the general principles concerning state responsibility in Customary International Law. These hold that in case of illegal acts,

> "...reparation must, as far as possible, wipe out all the consequences of the illegal act and re-establish the situation which would, in all probability, have existed if that act had not been committed"[11]

and

> "[t]he compensation shall cover any financially assessable damage including loss of profits insofar as it is established."[12]

Arbitration panels have assessed fair market value in a variety of ways, some forward-looking (such as discounted cash flows), and some backward-looking (incurred investment costs, for instance). Importantly for what follows, the purpose of the payment is to compensate the injured party for its losses, not to punish the responsible state: "A tribunal shall not award punitive damages."[13]

Finally, many investment agreements include a compulsory dispute settlement mechanism. These mechanisms differ from those in trade agreements in several fundamental respects. For instance, they do not only allow state-to-state disputes, but also allow foreign investors to litigate against host country governments (ISDS). Another difference is that the enforcement of rulings is much more potent than in trade agreements, since prevailing investors often can request courts at home, in the host country, as well as in third countries, to seize assets belonging to the host country in case the losing state does not willingly abide by the ruling.

As stated above, the purpose of the paper is to examine the design and implications of investment agreements constrained to share certain basic features with actual agreements. We therefore impose the following restrictions in most of the analysis to follow:

Contract Restrictions *Feasible agreements have the following features:*
(1) Investment decisions and regulation decisions are left at the discretion of investors and the host country, respectively.

[11] This often quoted passage is from the determination by the Permanent Court of International Justice in *The Factory at Chorzów* case 1928.
[12] Article 36, International Law Commission (2001). A footnote is omitted.
[13] Crawford (2002, p. 219).

(2) Agreements impose no taxation or performance requirement on investors;
(3) There are no payments to or from outside parties;
(4) Compensation is non-negative and paid only in case of regulation; and
(5) Compensation is proportional to, and not larger than, foregone operating profits.

These restrictions ensure that feasible agreements cannot subsidize investment, nor impose punitive damages on host countries, since compensation cannot exceed foregone operating profits and can only be paid in case of regulation. This does not preclude other legal arrangements between host countries and individual firms or the industry, but these would then be subsumed in the domestic welfare and profit functions. The main deviation from these restrictions occurs in Section 5.5, where we allow compensation to diverge from operating profits.

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

The most debatable constraint is Contract Restriction (5). From a contract point of view it might of course be better for the countries to invoke more general compensation schemes. We indeed examine such schemes in an extension to the main findings. In reality, contract stipulations usually relate compensation payments to notions of the market value of the investment and specifically to losses of profit in certain cases. In this respect, the design of investment agreements reflects the above-mentioned basic principles concerning the limits to state responsibility. A consistent interpretation of such provisions is to assume that compensation should be proportional to, but at most equal to, operating profit. Based on this set of restrictions, we will show that it is optimal to pay firms full compensation whenever they are entitled to damage payments.

Finally, we assume that agreements are perfectly enforceable. This seems reasonable considering the strong enforcement possibilities offered by the ISDS mechanisms. It can also be noted that perfect enforcement is a well-established assumption e.g. in the trade agreement literature, despite much weaker enforcement mechanisms in those agreements.

3 The setting

Consider an industry in country i consisting entirely of firms from country j. The interaction in the industry occurs in stages. The foreign firms first make simultaneous irreversible investments in production facilities. The host country is subsequently hit by a country-specific exogenous shock that affects the country's benefits from production. Having observed the shock, the host country decides whether to allow production, or to regulate by shutting down the industry. In the final stage, there is production and consumption unless the industry is regulated.

The two countries are interrelated in that the decision on regulation in country i affects the profits of investors from j. This is the direct effect on the value of the investment. We assume that regulatory decisions are strategically independent, however. This assumption simplifies that exposition, and allows us to focus on what we see as the most relevant aspects of investment agreements. For the sake of expositional convenience, the main text uses a standard model of a perfect competitive representative firm. Most of the results are derived in the Appendix using a significantly more general framework that admits imperfect competition, firm asymmetries, etc.

3.1 Product market competition

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

The total production cost of a representative foreign firm j is $C^j(x_i, k_i)$, where x_i is its production volume, and k_i its investment. The cost function has standard properties: $C^j(0, k_i) = 0$; $C^j_x > 0$; $C^j_{xx} > 0$; $C^j_k < 0$; $C^j_{kk} > 0$; $C^j_{xk} < 0$; and $C^j_{xx}C^j_{kk} \geq C^j_{xk}C^j_{kx}$.[14]

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

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

with $X^i_k(k_i) > 0$ and $P^i_k(k_i) < 0$; see Appendix A.1 for the derivations of these and other comparative statics. To distinguish between the maximization problem facing the price-taking investors, and the problem of maximizing aggregate welfare, we define two reduced form expressions for operating profits:

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

The total welfare that the host country derives from production in the industry—its "domestic welfare"—is

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

where the first two terms represent conventional consumer surplus (disregarding the constant income Υ^i). The last term is a production externality, the magnitude of which depends on a stochastic

[14] Subscripts attached to function operators denote partial derivatives.

shock $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$. The production externality Ψ^i can either be positive or negative, but a larger shock θ_i is defined always to correspond to a more negative externality: $\Psi^i_\theta < 0$. This negative effect dominates for sufficiently large shocks, in which case the host country will regulate absent an investment agreement. In contrast to much of the takings literature, we allow the externality also to be a function of the investment, which for simplicity is concave: $\Psi^i_{kk} \leq 0$.

High realizations of θ_i could represent the arrival of information concerning adverse environmental or health consequences of the production process, as in the case of the Fukushima disaster, or concerning product characteristics, as in the case of tobacco. The model could also capture a situation where a host country has made implicit or explicit promises to pursue a certain policy, but where a financial shock induces the country to change its policy, such as in the case of the subsidies to renewable energy in Spain. We will not adopt any particular interpretation, but simply denote it as a "regulatory shock."

Several comments on the model are in order. The model allows us to distinguish between pecuniary benefits from production stemming from local consumption, $\Omega^i(z_i) - p_i z_i$, and production externalities $\Psi^i(k_i, \theta_i)$. The model is also compatible with positive externalities from investment in the form of employment, technological spill-overs, learning-by-doing by the work-force, etc. Such effects would be subsumed in the expression for externalities $\Psi^i(k_i, \theta_i)$. For instance, we could assume that $\Psi^i(k_i, \theta_i) \equiv \Lambda^i(X^i(k_i)) - \Phi^i(X^i(k_i), \theta_i)$, where $\Lambda^i(X^i(k_i))$ captures the externalities for the local economy from foreign production, and $\Phi^i(X^i(k_i), \theta_i)$ the adverse effects of the regulatory problem that stochastically affects the economy ($\Phi^i_\theta > 0$).

Second, it is commonplace to refer to the benefits of foreign direct *investment* when pointing to the positive effects of foreign firms in the local economy. Some of the local employment and business effects certainly are directly related to the investment because they materialize during the construction of the plant. These specific externalities are sunk when the country decides whether to regulate and we therefore assume them to be zero. Instead, $\Psi^i(k_i, \theta_i)$ refers to production externalities that vanish in case of regulation. Since both positive and negative externalities are likely to increase in the investment k_i, we allow the marginal production externality to be positive or negative, and to depend on the shock: $\Psi^i_k(k_i, \theta_i) \gtrless 0$.

3.2 Regulation absent investment protection

Host country i observes θ_i and then decides whether to allow production or to regulate the industry. Domestic welfare equals $S^i(k_i, \theta_i)$ and foreign firms make the operating profits $\Pi^j(k_i)$ when production is allowed. By assumption, all externalities from the investment are related to production, so regulation by country i implies that $\Psi^i = 0$, there is no consumption, and operating profits are zero. Let domestic welfare be strictly positive for any $k_i > 0$ if the regulatory shock equals $\underline{\theta}_i$, but very negative if the shock equals $\bar{\theta}_i$. The regulatory shock $\Theta^i(k_i)$ for which the host country is indifferent between allowing production and regulating then is given by

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

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

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

$$S^i(k_i,\Theta^{iG}) + \Pi^j(k_i) \equiv 0.$$

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

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

3.3 Investment absent investment protection

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

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

The efficient level of investment maximizes

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

given $\hat{\theta}_i$. The associated first-order condition gives the efficient investment as a function $K^{iG}(\hat{\theta}_i)$ of the foreseen cut-off level for regulation $\hat{\theta}_i$:

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

A comparison of (3) and (5), using $K^i_\theta > 0$, yields

$$K^i(\hat{\theta}_i) < K^{iG}(\hat{\theta}_i) \text{ iff } \int_{\underline{\theta}_i}^{\hat{\theta}_i} \Psi^i_k(K^{iG}(\hat{\theta}_i),\theta_i)dF^i(\theta_i) > 0. \tag{6}$$

Hence:

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

3.4 The inefficiency of the outcome absent investment protection

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

$$\hat{\theta}_i < \Theta^i(K^i(\hat{\theta}_i)) \text{ iff } \hat{\theta}_i < \theta', \tag{7}$$

and we make the corresponding assumption regarding the function Θ^{iG}. These assumptions correspond to the "stability" conditions used for instance in oligopoly models to rule out counter-intuitive comparative statics properties.[15]

Condition (7) has the intuitively appealing implication that the direct reduction in investment that results from the host country's disregard of foreign investor interests in its regulatory decision, does not indirectly induce the host country to reduce its regulation to the extent that there is in equilibrium *less* regulation than there would be if the host country took full account of foreign investors' profits. The latter outcome (k_i^E, θ_i^E) would be given by $k_i^E = K^i(\theta_i^E)$ and $\theta_i^E = \Theta^{iG}(k_i^E)$:

Observation 3 *Absent investment protection, the host country regulates more frequently ($\theta_i^N < \theta_i^E$), and firms invest less ($k_i^N < k_i^E$), than when regulation is efficient.*[16]

Full (unconstrained) efficiency requires that investment and regulation maximize joint welfare. Denoting variables pertaining to such an outcome by superscript "G" (for "global"), the fully efficient solution is defined by $k_i^G = K^{iG}(\theta_i^G)$ and $\theta_i^G = \Theta^{iG}(k_i^G)$. We can decompose the total investment distortion $k_i^G - k_i^N$ absent any agreement into two parts. First, $k_i^E - k_i^N > 0$ is the

[15] The stability condition (7) implies that there exists at most one solution θ'. Existence follows by way of the Mean-Value Theorem, $\Theta^i(K^i(\underline{\theta}_i)) - \underline{\theta}_i \geq 0$ and $\Theta^i(K^i(\bar{\theta}_i)) - \bar{\theta}_i \leq 0$.

[16] If $\theta_i^E \leq \theta_i^N$, then $\theta_i^E \leq \Theta^i(k_i^E) < \Theta^{iG}(k_i^E) = \theta_i^E$, where assumption (7) implies the weak inequality and $\Theta^{iG}(k_i) > \Theta^i(k_i)$ the strict inequality. This is a contradiction, so $\theta_i^N < \theta_i^E$. Then, $k_i^N < k_i^E$ follows from $K^i_\theta > 0$.

distortion of investment stemming from the host country's disregard of foreign interests when regulating. The remaining part, $k_i^G - k_i^E \gtrless 0$, reflects the firms' neglect of production externalities. The sign of the latter is ambiguous since the marginal production externality can be positive or negative. Still, the model can generate overregulation and underinvestment also relative to the first best: $\theta_i^N < \theta_i^G$ and $k_i^N < k_i^G$. For expositional reasons we will think of this as the outcome absent an agreement, but none of the results hinge on this.

4 Investment agreements

As was shown above, the irreversibility of investment and the host country disregard of the interests of foreign investors can yield an outcome with overregulation and underinvestment. There is therefore potentially scope for an investment agreement that stimulates foreign investment by mitigating the host country incentive to regulate. A very simple agreement would request the host country to fully compensate investors for foregone operating profits whenever there is regulation. But full compensation would cause investors to attach weight also to realizations of θ_i for which their investments have no social value, and could thereby result in *excessive* investment. Simple as this observation is, it points to a fundamental feature of regulatory expropriations that proponents of investment agreements often seem to disregard: since regulations normally exist to address potential *overinvestment* problems, it is inherently possible for investment agreements to overshoot their targets. Indeed, an agreement that compensates for all regulations might even reduce joint welfare relative to the no-agreement situation.[17] Consequently, an optimal investment agreement might require exemptions from compensation. But how should such carve-outs optimally be designed, and what are the implications of such agreements?

4.1 Do agreements cause regulatory chill?

A core claim in the policy debate holds that investment agreements cause *regulatory chill*, although the notion is rarely precisely defined.[18] It can be given at least two different interpretations within the context of our model: We will say that an agreement causes *domestic* regulatory chill if the associated compensation scheme prevents host countries from undertaking regulations they would chose absent compensation. A corresponding *global* regulatory chill occurs if the agreement induces host countries to allow production in situations where regulation would be desirable from a joint welfare perspective. The following result, which we prove in Appendix A.2.3, has immediate implications for the existence of regulatory chill:

Proposition 1 *For any investment agreement satisfying Contract Restrictions (1)-(4) there exists an alternative agreement satisfying the same restrictions that for each host country i:*

[17] An example verifying this claim is available upon request. This particular moral hazard problem was pointed out by Blume et al (1984).

[18] Janeba (2016) is an exception.

(i) implements a threshold function for regulation $\Theta^{i*}(k_i) \in [\Theta^i(k_i), \Theta^{iG}(k_i)]$*; and*
(ii) yields weakly higher expected domestic welfare and foreign industry profits than under the initial agreement.

It follows directly from the Proposition that there will be no global regulatory chill. On the contrary, there will be regulation for a range of θ_i for which production would have been ex post efficient, i.e. $\theta_i \in (\Theta^{i*}(k_i), \Theta^{iG}(k_i))$. Hence:

Corollary 1 *An optimal investment agreement implies domestic, but not global, regulatory chill. The agreement will either induce ex post efficient, or excessive regulation from a joint welfare perspective.*

To see the generality of this finding, note first that global regulatory chill can only occur if the agreement stipulates compensation in excess of foregone operating profits for some $\theta_i > \Theta^{iG}(k_i)$. Firms earn $\Pi^j(k_i)$ for such realizations of θ_i, since they are allowed to produce in case of regulatory chill. Reducing compensation to $\Pi^j(k_i)$ would instead induce the host country to regulate. The investment incentives would remain unchanged because firms would still receive their operating profits for those shocks, but now as compensation for regulation. The modification of the compensation scheme thus increases regulatory efficiency by eliminating global regulatory chill without influencing investments or profits, and therefore represents a Pareto improvement.

Proposition 1 has several other noteworthy implications. First, an optimal agreement never induces regulation in instances where it would be efficient for the host country to allow production absent an agreement, i.e. for $\theta_i \leq \Theta^i(k_i)$. For interventions to be ex post optimal for the host country, firms would have to *pay* compensation when being regulated. This is ruled out by Contract Restriction (4), but a non-negativity constraint could also reflect limited liability. Second, optimal agreements feature domestic regulatory chill for a range of moderate regulatory shocks $\theta_i \in (\Theta^i(k_i); \Theta^{i*}(k_i))$. But domestic regulatory chill then simply reflects the fact that an agreement must dissuade the host country from regulating in certain circumstances for the agreement to be meaningful.

A third fundamental property of regulation in optimal investment agreements that is identified in Proposition 1 is the existence of a threshold value for regulation:

Corollary 2 *For any arbitrary investment* k_i*, it is ex post optimal for the host country to allow production if and only if the shock is below a threshold* $\Theta^{i*}(k_i)$*.*

It is of course always preferable from an ex post welfare point of view to regulate for very large shocks because the net welfare benefit of allowing production in host country i is strictly decreasing in θ_i. But the agreement is designed prior to the realization of the shock, so every possible outcome must be weighted by the density $f^i(\theta_i)$ to obtain the expected net benefit of allowing production. Hence, the ex ante optimal compensation scheme could in principle yield non-monotonic regulation in θ_i, depending on the properties of the density function $f^i(\theta_i)$ of the shock. Still, it is possible to hold the probability of regulation constant and to ensure production if and only if the shock is sufficiently mild, by awarding firms full compensation for all shocks below an appropriately chosen threshold. As demonstrated in the proof of the Proposition, this "reshuffling" of the probability of regulation increases regulatory efficiency and can be done without affecting firms' incentives to invest or their expected profits by a corresponding adjustment of the compensation payments.

The proof of Proposition 1 is constructive, showing how one can replace any initial compensation scheme that satisfies Contract Restrictions (1)-(4), with another scheme that satisfies the same restrictions, and that has the features listed in the Proposition. This alternative scheme uses a convex combination of the firm's foregone operating profit and the payment under the initial compensation scheme. 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. Hence, Proposition 1 is valid under tighter restrictions on feasible agreements than those imposed by Contract Restrictions (1)-(4). For instance, it holds for compensation schemes that pay out at most the foregone industry profit as compensation, and in particular for proportional compensation schemes (Contract Restriction (5)), and in the presence of non-discrimination clauses.

Note that Proposition 1 applies to a more general framework than the one laid out in Section 3. For instance, the setting employed in Appendix A.2.2 admits asymmetric firms, imperfect competition, and mixed foreign/domestic ownership structures, and the Proposition also holds when firms invest strategically to influence regulatory decisions.

4.2 How should investors be compensated?

Related to the issue of whether investment agreements cause regulatory chill are questions regarding when, and by how much, investors should be compensated in case host country policy measures significantly reduce their profits. To characterize an optimal compensation function we invoke also Contract Restriction (5), which stipulates that any compensation must be proportional to, and not larger than, foregone operating profits:[19]

$$\hat{T}^i(k_i, \theta_i) \equiv b_i(\theta_i)\Pi^j(k_i),\ b_i(\theta_i) \in [0, 1]. \tag{8}$$

The following Proposition characterizes optimal compensation schemes (the proof is provided in Appendix A.4):

Proposition 2 *For any investment agreement satisfying Contract Restrictions (1)-(5) there exists an alternative agreement satisfying the same restrictions, and that for each host country i:*

[19] An even more general proportional compensation mechanism would be $\hat{T}^i(k_i, \theta_i) \equiv b_i(k_i, \theta_i)\Pi^j(k_i)$. However, this formulation is equivalent to a non-linear compensation mechanism $\tilde{T}^i(k_i, \theta_i)$, as can be seen by letting $b_i(k_i, \theta_i) \equiv \tilde{T}^i(k_i, \theta_i)/\Pi^j(k_i)$. We consider non-linear compensation schemes in Section 5.5.

(i) features the compensation function

$$T^i(k_i, \theta_i) \equiv \begin{cases} \Pi^j(k_i) & \text{if } \theta_i \leq \hat{\theta}_i \\ 0 & \text{if } \theta_i > \hat{\theta}_i; \end{cases} \tag{9}$$

(ii) yields weakly higher expected domestic welfare and foreign industry profits than under the initial agreement.

Hence, the optimal compensation can be characterized entirely in terms of a simple *carve-out* from a full compensation requirement, whereby firms receive full compensation for foregone operating profits when regulation occurs for θ_i below a threshold value $\hat{\theta}_i$, and no compensation otherwise. We will refer to this as the *level of investment protection* in country i.

Again, Proposition 2 is more general than our setting above would suggest. A carve-out scheme is optimal also if we allow investors to take into account the effects of their investments on the probability of regulation; see Appendix A.3. A carve-out scheme is also optimal in the case of a strategic monopolist, in which case we can relax Contract Restriction (5) by assuming that compensation is non-negative but at most equal to foregone operating profit $\Pi^j(k_i)$; see Appendix A.5. The main difference from the competitive setting is that the threshold for compensation then depends on the level of investment.

Observe that while Contract Restriction (5) allows compensation to be strictly less than foregone operating profits, Proposition 2 nevertheless establishes the optimality of full compensation whenever firms are eligible for compensation. It will be shown below that full compensation can be sufficient to implement the fully efficient outcome (θ_i^G, k_i^G). Consequently, full compensation not only reflects a fundamental principle in Customary International Law, it also has desirable economic properties, by effectively inducing host countries to internalize all ramifications of their regulatory decisions:

Observation 4 *Compensation equal to foregone operating profits aligns host country incentives with the ex post efficient level of regulation. Under certain conditions, such compensation can be sufficient to correct both investment and regulation decisions.*

4.3 Can agreements solve the overregulation/underinvestment problems?

There is an ongoing policy debate concerning the need to redesign investment protection agreements, and is becoming increasingly common that model agreements are modified in order to reduce their ambits; this is being done by the EU and India, for instance. These developments can be given many explanations, but at a general level seem to reflect a dissatisfaction with the ability of the agreements to increase investment at an acceptable cost in terms of associated constraints on domestic policies. Such perceived shortcomings could in turn stem from a failure of the agreements to resolve overregulation/underinvestment problems in an efficient manner, or from the distribution of gains they produce. This Section will therefore investigate the extent to which agreement can resolve these inefficiencies, and the next Section will examine their distributional implications.

We here consider agreements designed to maximize the joint welfare of the parties. This analysis is of interest because it informs us about the maximal capacity of investment agreements to resolve the problems they are meant to address. More specifically, to what extent do Contract Restrictions (1)-(5) constrain the possible outcome away from the first-best outcome (θ_i^G, k_i^G)? But this constrained efficient outcome could also be a reasonable prediction for the equilibrium agreement in situations where the parties either are highly symmetric, or where they have access to some form of side payments.

To evaluate the efficiency of any agreement, we need to first derive formal expressions for the welfare of the parties to an agreement based on a threshold $\hat{\theta}_i$ for a host country i. The following feature is established in Appendix A.3:

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

It would only be possible to implement regulation for $\theta_i < \theta_i^N$ if the industry could be requested to *pay* compensation in case of regulation, but this would violate Contract Restriction (4). Any undertaking $\hat{\theta}_i < \theta_i^N$ thus is meaningless. And an undertaking $\hat{\theta}_i = \theta_i^N$ creates no strict gains for either party, and thus will not be formed either. Hence, any meaningful agreement necessarily features $\hat{\theta}_i > \theta_i^N$.

The expected domestic welfare of host country i is

$$\tilde{S}^i(\hat{\theta}_i) \equiv \begin{cases} \int_{\underline{\theta}_i}^{\hat{\theta}_i} S^i(\hat{k}_i, \theta_i) dF^i(\theta_i) \text{ for } \hat{\theta}_i \in [\theta_i^N, \theta_i^E) \\ \int_{\underline{\theta}_i}^{\Theta^{iG}(\hat{k}_i)} S^i(\hat{k}_i, \theta_i) dF^i(\theta_i) - [F^i(\hat{\theta}_i) - F^i(\Theta^{iG}(\hat{k}_i))]\Pi^j(\hat{k}_i) \text{ for } \hat{\theta}_i \in [\theta_i^E, \bar{\theta}_i], \end{cases} \tag{10}$$

where $\hat{k}_i = K^i(\hat{\theta}_i)$ is the equilibrium investment. The first row pertains to agreements with moderate investment protection, $\hat{\theta}_i \in [\theta_i^N, \theta_i^E)$, and the second row to the case of strong investment protection, $\hat{\theta}_i \in [\theta_i^E, \bar{\theta}_i]$. Hence, compensation function (9) effectively induces the host country to internalize the joint welfare effects of regulation for shocks below $\hat{\theta}_i$, because the country then has to pay full compensation to foreign investors, but not for shocks above $\hat{\theta}_i$. Furthermore, it is jointly welfare optimal to allow production for all $\theta_i \leq \Theta^{iG}(\hat{k}_i)$, whereas regulation maximizes domestic welfare if $\theta_i > \Theta^i(\hat{k}_i)$. Note in particular the occurrence of overregulation for $\theta_i \in (\hat{\theta}_i, \Theta^{iG}(\hat{k}_i))$ under moderate investment protection, whereas all regulation is ex post efficient under a strong level of investment protection.[20]

[20]Stability condition (7) implies $\hat{\theta}_i \in [\Theta^i(\hat{k}_i), \Theta^{iG}(\hat{k}_i))$ for all $\hat{\theta}_i \in [\theta_i^N, \theta_i^E)$, but $\hat{\theta}_i \geq \Theta^{iG}(\hat{k}_i)$ for all $\hat{\theta}_i \in [\theta_i^E, \bar{\theta}_i]$. Under moderate investment protection, the host country allows production for all $\theta_i \leq \hat{\theta}_i < \Theta^{iG}(\hat{k}_i)$, and regulates for all $\theta_i > \hat{\theta}_i \geq \Theta^i(\hat{k}_i)$. Under strong investment protection, the host country allows production for all $\theta_i \leq \Theta^{iG}(\hat{k}_i) \leq \hat{\theta}_i$, but regulates with compensation for $\theta_i \in (\hat{\theta}_i, \Theta^{iG}(\hat{k}_i)]$. Regulation maximizes domestic welfare also for all $\theta_i > \hat{\theta}_i > \Theta^{iG}(\hat{k}_i)$, but then the host country does not have to pay any compensation.

The expected welfare of source country j from its investments in country i equals the expected profits

$$\tilde{\Pi}^j(\hat{\theta}_i) \equiv F^i(\hat{\theta}_i)\Pi^j(\hat{k}_i) - R^j(\hat{k}_i). \tag{11}$$

For $\theta_i \leq \hat{\theta}_i$ it is immaterial to foreign investors whether they are allowed to produce and earn their operating profits, or whether the industry is regulated and the investors make the same profits in the form of compensation payments. For $\theta_i > \hat{\theta}_i \geq \Theta^i(\hat{k}_i)$ the industry is always regulated, but the host country does not have to pay any compensation. We assume throughout that the industry profit is increasing in the level of investment protection:

$$\tilde{\Pi}^j_\theta(\hat{\theta}_i) = \Pi^j(\hat{k}_i)f^i(\hat{\theta}_i) + X^i(\hat{k}_i)P^i_k(\hat{k}_i)K^i_\theta(\hat{\theta}_i)F^i(\hat{\theta}_i) > 0. \tag{12}$$

Intuitively, this assumption requires that the direct effect of improved investment protection dominates the indirect price effect of increased investments at the aggregate industry level.

Let the joint expected welfare be denoted $\tilde{W}^i(\hat{\theta}_i) \equiv \tilde{S}^i(\hat{\theta}_i) + \tilde{\Pi}^j(\hat{\theta}_i)$; we assume for simplicity that $\tilde{W}^i(\hat{\theta}_i)$ is strictly quasi-concave for $\hat{\theta}_i > \theta^E_i$. Also, let $\theta^W_i \equiv \arg\max_{\hat{\theta}_i \geq \theta^N_i} \tilde{W}^i(\hat{\theta}_i)$, with the corresponding investment level denoted $k^W_i \equiv K^i(\theta^W_i)$.

There are three qualitatively different types of outcomes, depending on the effect of increased investment on the production externality. A first case occurs when the marginal production externality evaluated at $\hat{\theta}_i = \theta^E_i$ is negative:[21]

$$\tilde{W}^i_\theta(\theta^E_i) = \int_{\underline{\theta}_i}^{\theta^E_i} \Psi^i_k(k^E_i, \theta_i)dF^i(\theta_i)K^i_\theta(\theta^E_i) < 0. \tag{13}$$

An increase in investment protection above θ^E_i would then only serve to exacerbate an overinvestment problem without improving the ex post incentive to regulate. The *constrained efficient* level of protection is the solution to $\tilde{W}^i_\theta(\theta^W_i) = 0$, and has the feature that $\theta^W_i < \theta^E_i$, with $k^W_i < k^E_i$. It strikes a balance

$$\Pi^j(k^W_i)f^i(\theta^W_i) = -\int_{\underline{\theta}_i}^{\theta^W_i} \Psi^i_k(k^W_i, \theta_i)dF^i(\theta_i)K^i_\theta(\theta^W_i) - S^i(k^W_i, \theta^W_i)f^i(\theta^W_i)$$

between the direct marginal benefit to foreign investors from investment protection (the term on the left-hand side of the above equation) against the worsened expected production externality (the first term on the right-hand side) plus the cost of additional domestic underregulation (the second term on the right-hand side).

In a situation with underinvestment absent an agreement, $k^N_i \leq k^G_i$, it would be possible to implement the efficient investment level k^G_i by an appropriate choice of investment protection $\theta'_i \in [\theta^N_i, \theta^E_i]$, but there would still be overregulation, $\theta'_i < \theta^G_i$.[22] Hence, it is impossible to obtain the fully

[21] Recall that the threshold $\theta^E_i = \Theta^{iG}(K^i(\theta^E_i))$ represents the equilibrium absent an agreement if the host country were to fully internalize all effects on foreign profit in its regulatory decision, but firms invest to maximize expected profit.

[22] The assumption of $\tilde{W}^i_\theta(\theta^E_i) < 0$ and strict quasi-concavity imply $k^G_i < k^E_i$. By the additional assumption $k^N_i \leq k^G_i$, we obtain a string of inequalities $K^i(\theta^N_i) = k^N_i \leq k^G_i < k^E_i = K^i(\theta^E_i)$. The mean-value theorem and monotonicity of K^i therefore imply the existence of a unique $\theta'_i \in [\theta^N_i, \theta^E_i)$ such that $K^i(\theta'_i) = k^G_i$. However, by the stability condition (7), it follows that $\theta'_i < \Theta^{iG}(K^i(\theta'_i)) = \Theta^{iG}(k^G_i) = \theta^G_i$.

efficient outcome (θ_i^G, k_i^G) with the carve-out scheme (9) when the marginal expected production externality as defined in (13) is negative.

The second case is where the marginal production externality is non-negative at θ_i^E, $\tilde{W}_\theta^i(\theta_i^E) \geq 0$, but not larger than

$$\tilde{W}_\theta^i(\bar{\theta}_i) = [\int_{\underline{\theta}_i}^{\Theta^{iG}(\bar{k}_i)} \Psi_k^i(\bar{k}_i, \theta_i) dF^i(\theta) - (1 - F(\Theta^{iG}(\bar{k}_i)) R_k(\bar{k}_i)] K_\theta^i(\bar{\theta}_i) \leq 0, \tag{14}$$

where $\bar{k}_i = K^i(\bar{\theta}_i)$ is the investment level that results if foreign investors are fully compensated for all regulation. In this case, the carve-out scheme (9) is sufficient to implement the fully efficient outcome (θ_i^G, k_i^G). Increasing the level of investment protection $\hat{\theta}_i$ up to the level θ_i^E improves firms' ex ante incentives to invest, and simultaneously removes the host country's incentive to overregulate. Regulation is ex post efficient for strong protection levels $\hat{\theta}_i \geq \theta_i^E$, and $\hat{\theta}_i$ can then be used exclusively to improve the industry's investment incentives.

In the third case, the marginal expected production externality is too large for it to be possible to implement the fully efficient investment level even under complete investment protection: $\tilde{W}_\theta^i(\bar{\theta}_i) > 0$, and $k_i^G > \bar{k}_i$. In this situation, the welfare maximizing choice offers full protection of operating profits, $\theta_i^W = \bar{\theta}_i$, but there will still be underinvestment.

We summarize the above findings regarding an agreement that maximizes joint welfare as follows (see Appendix A.5 for a formal proof):

Proposition 3 *A constrained efficient investment agreement satisfying Contract Restrictions (1)-(5) achieves the fully efficient outcome (θ_i^G, k_i^G) when the marginal expected production externality is positive, but not too large. In other situations the agreement can only achieve second-best outcomes.*

To conclude, it is hardly surprising that Contract Restrictions (1)-(5) hinder the agreement from implementing the first best in some situations. More surprising is the fact that it is still possible to correct both the regulatory and the investment distortion by means of the single policy variable $\hat{\theta}_i$ in a robust set of circumstances.[23,24]

4.4 Who benefits and who loses?

We now turn to distributional aspects of optimal investment agreements. The equilibrium level of investment protection is determined through negotiations between the parties to an agreement. One

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

[24] We conjecture that a carve-out policy is sufficient to maximize joint welfare also for the negative marginal production externality (13) if compensation $\hat{T}_i(k_i, \theta_i) \geq 0$ is required to be non-decreasing in k_i. It should then be optimal to set $\hat{T}_k^i(k_i, \theta_i) = 0$ to reduce firms' overinvestment. The remaining problem would be to maximize joint welfare over $\hat{\theta}_i$. Non-negative marginal compensation, $\hat{T}^i(k_i, \theta_i) \geq 0$, would arise under a restriction that all compensation be equal to the fair market value of the investment and if this value is non-decreasing in the amount of capital invested.

would therefore expect the equilibrium investment protection to differ from the constrained efficient level θ_i^W, but an agreement can generally speaking result in either over- and underinvestment.[25] Source country j would prefer full protection for its investment in country i, $\hat{\theta}_i = \bar{\theta}_i$, whereas country i would prefer the level of investment protection that maximizes its expected domestic welfare, $\theta_i^U \equiv \arg\max_{\hat{\theta}_i \geq \theta_i^N} \tilde{S}^i(\hat{\theta}_i)$. Consequently, the countries have a common interest in setting $\hat{\theta}_i$ no lower than θ_i^U. The assumption that the parties split the surplus from the agreement thus implies the following:[26]

Observation 5 *The negotiated level of investment protection, and the level of investment, will in any agreement exceed the level that maximizes host country domestic welfare: $\hat{\theta}_i \geq \theta_i^U$, with strict inequality unless the host country prefers maximal investment protection ($\theta_i^U = \bar{\theta}_i$).*

The negotiated level of protection will also depend on whether the agreement covers one-way or two-ways flows of investment. When investments flow in both directions, it is possible to negotiate exchanges of concessions across industries. In the special case of agreements between fairly symmetric countries, such as the TTIP negotiations between the EU and the US, it would be natural to assume that the negotiated level of investment protection $\hat{\theta}_i$ de facto maximizes the joint expected welfare.

To obtain sharper predictions concerning the distributional effects of agreements, we will contrast two stylized types of agreements. The first scenario is the agreement between a developed and a developing country. While often formally symmetric, such agreements are in practice highly asymmetric since they are meant to encourage increased investment flows from developed to the developing countries only. We denote this a "North-South" agreement. The other scenario is an agreement between two developed countries. Such agreements typically differ from North-South agreements in two respects. First, there are investment flows in both directions between the contracting parties. Second, countries are able to make credible unilateral commitments with regard to investment protection absent any agreement. To simplify matters we will focus on situations where the countries are mirror images in terms of demand structures, technologies, and propensity to experience regulatory shocks; this also seems broadly descriptive of the conditions that faced e.g. the EU-US negotiations over TTIP. We will refer to this as a "North-North" agreement. We use

[25] A constrained efficient agreement could of course be achieved by pure coincidence, for instance under a specific distribution of the bargaining strength. Assume that the negotiated investment protection level $\hat{\theta}_i$ maximizes the Nash product $[\tilde{S}^i(\hat{\theta}_i) - \tilde{S}^i(\theta_i^N)]^{\rho_i}[\tilde{\Pi}^j(\hat{\theta}_i) - \tilde{\Pi}^j(\theta_i^N)]^{\rho_j}$, where ρ_i (ρ_j) is the bargaining strength of the host (source) country. Using $\tilde{S}_\theta^i(\theta_i^W) + \tilde{\Pi}_\theta^j(\theta_i^W) = 0$, it is straightforward to verify that the negotiated solution equals θ_i^W if

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

[26] A marginal increase in $\hat{\theta}_i$ at θ_i^U has a first-order effect on industry profit, $\tilde{\Pi}_\theta^j(\theta_i^U) > 0$, but only a second-order effect on domestic expected welfare, $\tilde{S}_\theta^i(\theta_i^U) = 0$, if $\theta_i^U \in (\theta_i^N, \bar{\theta}_i)$. Setting $\hat{\theta}_i > \theta_i^U$ increases thus total surplus. If $\theta_i^U = \theta_i^N$, Lemma 1 implies that an investment agreement is economically meaningful only if $\hat{\theta}_i > \theta_i^N = \theta_i^U$.

throughout the situation absent an agreement as a benchmark for evaluating the consequences of the agreements.

The expected benefit to North (country j) of a North-South agreement is $\tilde{\Pi}^j(\hat{\theta}_i) - \tilde{\Pi}^j(\theta_i^N)$. Since the expected industry profit increases with higher levels of investment protection, North would benefit from any agreement with $\hat{\theta}_i > \theta_i^N$. A necessary and sufficient condition for there to be scope for a North-South agreement is therefore that South unilaterally prefers a protection level $\theta_i^U > \theta_i^N$. If markets are perfectly competitive, there are no first order effects from an agreement in terms of induced changes in production or investment levels. Instead, South benefits from an agreement if the combined effect of the fall in the price of the good and the expected change in the production externality is positive:

$$\tilde{S}_\theta^i(\theta_i^N) = -P_k^i(k_i^N)X^i(k_i^N)F^i(\theta_i^N) + \int_{\underline{\theta}_i}^{\theta_i^N} \Psi_k^i(k_i^N, \theta_i)K_\theta^i(\theta_i^N)dF^i(\theta_i) > 0. \tag{15}$$

In fact, South could benefit from an agreement even if the expected production externalities were negative, provided the benefits of the price reductions were sufficiently large. South's motive for entering into an agreement with North would then be pure rent-shifting. But the agreement would still benefit North since it would increase the range of θ_i for which there is no regulation; recall our assumption that the direct benefit of investment protection dominates the negative price effect in (12). However, it is entirely possible that the expected production externality is sufficiently negative to outweigh any positive price effects, so that $\theta_i^U = \theta_i^N$, even if an agreement would increase joint welfare:

$$\tilde{W}_\theta^i(\theta_i^N) = \int_{\underline{\theta}_i}^{\theta_i^N} \Psi_k^i(k_i^N, \theta_i)dF^i(\theta_i)K_\theta^i(\theta_i^N) + \Pi^j(k_i^N)f^i(\theta_i^N) > 0. \tag{16}$$

There are thus two first-order divergences between host country interests and the aggregate effects in a North-South setting. First, the reduction in the consumer price that follows from increased investment protection —the first term in (15)—is only a transfer from firms to consumers from a global perspective. Second, the host country does not take into consideration the expected increase in operating profits of foreign investors from less frequent regulation—the second term in (16). An investment agreement might thus fail to form despite being desirable from a global perspective.[27]

In the case of a North-North agreement the expected welfare of country i equals $\tilde{S}^i(\theta_i^W)+\tilde{\Pi}^i(\theta_j^W)$, where $\theta_i^W = \theta_j^W$ by the assumed symmetry. As countries can unilaterally commit with regard to compensation, country i will not accept any agreement that yields less than $\tilde{S}^i(\theta_i^U) + \tilde{\Pi}^i(\theta_j^U)$, where $\theta_i^U = \theta_j^U$ due to the assumed symmetry. Also, $\tilde{\Pi}^i(\theta_j^W) = \tilde{\Pi}^j(\theta_i^W)$ and $\tilde{\Pi}^i(\theta_j^U) = \tilde{\Pi}^j(\theta_i^U)$ by the symmetry, so any agreement satisfies $\tilde{W}^i(\theta_i^W) > \tilde{W}^i(\theta_i^U)$ for both countries. Hence, a North-North agreement will arise if and only if it increases joint welfare. A sufficient condition for this is $\tilde{W}_\theta^i(\theta^N) > 0$ even if $\theta^U = \theta^N$.[28] But the gains from a North-North agreement come with

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

[28] One could argue that Northern countries should not be constrained to compensation schemes fulfilling Contract Restrictions (1)-(5) when making unilateral commitments to investment protection. If so, this would reduce the scope

pronounced distributional implications. Since country i can unilaterally implement its preferred protection level for inflowing FDI, it will suffer the loss $\tilde{S}^i(\theta_i^U) - \tilde{S}^i(\theta_i^W) > 0$ in domestic welfare by signing an agreement with j. Country i will therefore accept such an agreement if and only if it is compensated for this domestic loss in terms of better investment protection abroad for the FDI of its domestically-owned industry: $\theta_j^W > \theta_j^U$ is sufficiently large that $\tilde{\Pi}^i(\theta_j^W) - \tilde{\Pi}^i(\theta_j^U) > \tilde{S}^i(\theta_i^U) - \tilde{S}^i(\theta_i^W)$.

The above arguments and Observation 5 jointly imply:

Proposition 4 *A North-South agreement might not form even if it increases expected joint welfare. But if it does, the agreement benefits foreign investors and increases domestic welfare in South. A North-North agreement is formed if and only if it increases expected joint welfare. Such an agreement benefits foreign investors in both countries, but reduces expected domestic welfare in both countries.*

We believe that these observations shed light on the political economy of 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 have been 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. Our North-South and North-North scenarios point to a possible explanation for the much more critical public view of the mega-regional agreements: the existing legal systems in e.g. the EU and the U.S. already provide sufficient protection of FDI to internalize all domestic welfare effects. The *additional* investment protection that an agreement such as TTIP would offer would mainly benefit foreign investors, but would 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.

4.5 Two different roles of investment agreements

The stylized North-South and North-North scenarios identify two different roles that investment agreement might play. The welfare benefits from a North-South investment agreement stem entirely from the credibility the agreement lends to South's commitment to compensate for regulation for a range of $\theta_i > \theta_i^N$. If South had full unilateral commitment possibilities, it would choose θ_i^U absent any agreement, in which case there would be no gains from an agreement: $\tilde{S}^i(\hat{\theta}_i) - \tilde{S}^i(\theta_i^U) \leq 0$ for all $\hat{\theta}_i$. But lacking this commitment ability, South has to negotiate with North over the level of investment protection, and will consequently have to accept a higher level of investment protection

for an investment agreement. But since host countries do not internalize the effects of their protection on foreign investors, there could still be room for a welfare improving agreement of the type we are considering.

than what is optimal from the point of view of domestic welfare. 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. 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.[29]

The benefits of a North-North agreement do not arise from this type of commitment, as countries are able to unilaterally commit to protecting incoming investment, but from *internalization of negative international externalities* from national regulatory policies. Since investments flow 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 6 *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 7 *Contracting on investment and regulation with individual investors directly would be better than a North-South agreement (absent transaction costs), but would not necessarily dominate a North-North agreement.*

4.6 Compensation for political vs. regulatory shocks

A fundamental principle in international law is that legitimate governments can make long-term commitments on behalf of their states. But it is often held that investment agreements impose undemocratic constraints on future governments. Other observers argue instead that investors should be protected against changes in political preferences, and that they should be compensated in case of regulation regardless of whether motivated by exogenous factors or changes in the political

[29] 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.

situation. To shed some light on this issue, we will distinguish between shocks that influence regulatory decisions for given political preferences, and changes in political preferences that affect regulatory decisions for given regulatory environments. An actual example of this is the acceleration of the German nuclear phase-out that was decided in the aftermath of the Fukushima disaster. This decision can either be viewed as motivated by the arrival of new information concerning the dangers of nuclear power, or as a political move by government facing dwindling popular support.[30] Both types of shocks have triggered litigation under investment agreements. Should they be treated differently?

To capture the distinction between exogenous and political risks, we make a very simple reformulation of the model by assuming that there are two shocks in country i, an exogenous regulatory shock η_i, and a political preference shock λ_i. The shocks are multiplicative such that the externality perceived by a government with the preference parameter λ_i is $\Psi^i(k_i, \theta_i)$, where $\theta_i \equiv \eta_i \lambda_i$. By this characterization, a higher λ_i implies a government that is more sensitive to exogenous information concerning the value of allowing production and therefore is more prone to regulate. Assume that both shocks are resolved simultaneously between the investment stage and the production stage, and let $F^i(\theta_i)$ be the cumulative distribution of the total shock in host country i. This reformulated version of the model is formally identical to the one above, except with regard to the interpretation of θ_i.

Assume that the agreement is negotiated behind a veil of ignorance and designed to maximize expected joint welfare, where the expectation is taken over both the regulatory and the political shock. In this case, Proposition 1 still characterizes an optimal investment agreement as a function of $\hat{\theta}_i$, and θ_i^W is the (constrained) jointly welfare maximizing level of investment protection. By assumption, the two shocks affect the outcome in a completely symmetric fashion and therefore are perfect substitutes. The corresponding welfare maximizing level of investment protection for exogenous regulatory shocks η_i thus equals $\eta_i^W = \theta_i^W / \lambda_i$ for any political preference λ_i. Hence:

Observation 8 *The compensation to firms should depend on political preferences in a welfare optimal investment agreement. With the total shock given by $\theta_i \equiv \eta_i \lambda_i$, a government that is more sensitive to exogenous shocks (λ_i is higher) should be allowed to offer less investment protection for such shocks (η_i^W is smaller).*

We have here used a highly reduced form of political preference representation. However, the multiplicative specification was only chosen for expositional simplification. More critical is the assumption that the compensation scheme is designed behind the veil of ignorance. The properties of the agreement are likely to reflect the political preferences of the government in place when the agreement is negotiated. The current government can then bind the regulatory decisions of future governments independently of future political preferences by establishing that regulation

[30] Another example of political changes triggering expropriation, is the wave of nationalization and regulation that occurred in Venezuela after Mr. Hugo Chavez came to power, and that led to a number of litigations.

should depend entirely on η_i instead of θ_i. Hence, political preference shocks could be left out of investment agreements for political economy reasons, even if it would have been efficient to include them.

5 Extensions

This section extends the above analysis in a number of directions. We first consider non-discrimination and direct expropriation clauses. Then we turn to the robustness of the optimal compensation rules, to assumptions concerning market structure, and to asymmetric information. Finally, we derive a new compensation scheme based on relative performance that can implement the global optimum under a range of circumstances. For simplicity, we assume that countries are completely symmetric, and we suppress country indices throughout.

5.1 A National Treatment clause

Investment agreements typically include non-discrimination clauses. We have so far steered away from discrimination issues simply by assuming that there are no domestic firms. This is not quite as restrictive as it might seem, since there are in practice many instances where a National Treatment (NT) clause will not have any bite due to a lack of domestic firms that operate under sufficiently "like circumstances" to those facing foreign investors. To shed some light on the role of NT in our setting, we now assume that host countries feature 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 regulation decisions. Each host country fully internalizes the consequences of regulation for the profits of its domestic industry, but continues to disregard the impact on foreign profits.

To identify the nature of discrimination, note first that the host country regulates the foreign industry more frequently than the domestic industry if there is no investment protection: $\theta_F^N = \theta^N$ and $\theta_D^N = \theta^E > \theta^N$, with the corresponding investment levels $k_F^N = k^N$ and $k_D^N = k^E > k^N$; see Observation 3. The host country chooses different levels of protection in the two sectors also if it is able to unilaterally commit to investment protection: $\theta_F^U = \theta^U$ and $\theta_D^U = \theta^W \geq \theta^U$, followed by equilibrium investments $k_F^U = K(\theta^U)$ and $k_D^U = K(\theta^W) \geq k^U$; these inequalities are strict in the standard case of incomplete investment protection ($\theta^U < \bar{\theta}$).

Foreign firms can be said to face two forms of discrimination in these circumstances. First, the host country regulates foreign investment more frequently, $\theta_F^N < \theta_D^N$ and $\theta_F^U \leq \theta_D^U$, despite both industries being subject to the same shock θ. Second, the host country applies a stricter rule for when to regulate industry F for any investment level k, since $\Theta(k) < \Theta^G(k)$. 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.

5.1.1 An agreement comprising NT only

A common view among critics of investment agreements is 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." Imposing an NT provision in a no-agreement situation would force investment protection to a common level $\theta_F = \theta_D$ because there would be more frequent regulation of foreign investments absent an agreement, regardless of whether the host country can commit unilaterally or not. Since foreign and domestic firms are fully symmetric, they will make the same investment under an NT clause.

In a North-South agreement on NT only, the unilaterally determined common investment protection level θ^{NS}_{NTOnly} would be given by

$$2S(K(\theta^{NS}_{NTOnly}), \theta^{NS}_{NTOnly}) + \Pi(K(\theta^{NS}_{NTOnly})) \equiv 0$$

where 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 $\theta^{NS}_{NTOnly} \in (\theta^N_F, \theta^N_D)$ under a similar stability condition as (7). Consequently, there will be increased foreign investment, and reduced domestic investment. This benefits North, but has ambiguous implications for domestic welfare in South.

In a North-North agreement on NT only, the unilaterally determined common protection level θ^{NN}_{NTOnly} has the following features (see Appendix A.4 for a proof):

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

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. By way of the two-way investment flows, both countries benefit from the better treatment of their foreign investment, and it is ambiguous whether they would accept such an agreement. Indeed, in the present setting it might actually be better from a joint welfare perspective to have no agreement at all compared to an agreement that only imposes NT.

Proposition 5 *The imposition of only an NT provision:*
(i) has ambiguous welfare consequences in a North-North agreement;
(ii) benefits North, and has ambiguous welfare consequences for South, in a North-South agreement.

More could be achieved if the parties also negotiated investment protection levels for the foreign industries. In particular, a more extensive agreement would facilitate an internalization of international externalities from domestic regulation in case of two-way investment flows.

5.1.2 NT as a complement to undertakings on investment protection levels

Investment agreements usually include both NT and investment protection obligations. How does the presence of NT affect the optimal design of the latter provisions? A North-North agreement will yield the same efficient protection level in the 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: $\theta_D = \theta_F = \theta^W$. Consequently, there is no role for NT here.

The picture is more complex in the North-South case. An NT clause is again redundant if the agreement without NT already implies an investment protection level θ^{NS} for the foreign industry that exceeds the level θ^E for the domestic industry. Hence, NT has an impact only if South has sufficient bargaining power absent NT to achieve $\theta^{NS} < \theta^E$. Let θ^{NS}_{NT} be the negotiated level of investment protection in a Pareto optimal North-South agreement under NT. We prove in Appendix A.4 that an NT clause in a North-South agreement increases the level of investment protection above the level of a North-North agreement comprising only NT:

Lemma 3 $\theta^{NS}_{NT} \geq \theta^{NN}_{NTOnly}$*, with strict inequality if* $\theta^{NN}_{NTOnly} \in (\theta^N, \bar{\theta})$*.*

If South has most of the bargaining power, the negotiated outcome will be close to θ^U absent NT. In particular, $\theta^U < \theta^{NS} \leq \theta^{NN}_{NTOnly} < \theta^{NS}_{NT}$ if also $\theta^{NN}_{NTOnly} \in (\theta^N, \bar{\theta})$. North is then strictly better off with an NT provision included in the agreement, since this will increase the level of investment protection ($\theta^{NS}_{NT} > \theta^{NS}$). The effect on South is ambiguous. The domestic 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, its 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 6 *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.

NT is normally perceived as an instrument that prevents parties to an agreement from exploiting contractual incompleteness and thereby undermine bargaining concessions. In the case of trade agreements, a myriad of domestic policies can be used to offset commitments concerning border instruments; for instance, a tariff binding could easily be rendered useless by the introduction of a "sales tax" solely levied on the imported product. 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.[31] The purpose of NT in the present context is different as it is not meant to neutralize opportunistic behavior—there is no commitment that is 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 9 *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.*[32]

Noted 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. We leave this issue for future research.

5.2 A direct expropriation clause

A primary objective of investment agreements is of course to prevent direct expropriation. Such instances are nowadays less common than during the 1960s and 1970s, but are not completely something of the past, as shown by Hajzler (2012). Investment agreements typically have stricter rules regarding compensation for direct than for indirect expropriation, in the sense that carve-outs from the expropriation clauses, which we described in Section 2, only apply to direct expropriation. This stricter attitude might seem intuitively appealing since direct expropriations are (at best) pure transfers of rents. Matters are not quite as simple from a contractual point of view, however. First, regulatory expropriations shut down production in the regulated entities, whereas production

[31] Horn (2006) examines the pros and cons of NT in trade agreements from this perspective.

[32] In the case of two-way investment flows and non-commitment such as in a South-South scenario, an efficient negotiation over investment protection for foreign investment absent NT would lead to the protection level θ^W in the F sectors in the two countries. Assuming that countries cannot unilaterally commit to investment protection for its domestic industry, they will apply a different threshold $\theta_D^N = \theta^E \neq \theta^W$ for regulation in the domestically-owned industry D. Adding NT will still be inconsequential if the marginal production externality as defined in Proposition 3 is non-negative, because then the foreign sector would enjoy (weakly) more protection than the domestic sector under the initial agreement: $\theta^W \geq \theta^E$. However, in the opposite case of a negative marginal expected production externality, the NT clause would enable countries to credibly reduce underregulation in their domestic sectors from θ^E to θ^W. This reduction in the level of investment protection represents a domestic welfare improvement that hurts the domestic industry less than it benefits the rest of the domestic economy: the protection level θ^W maximizes $\tilde{W}(\hat{\theta})$, and $\theta^E > \theta^W$ implies $\tilde{\Pi}(\theta^E) > \tilde{\Pi}(\theta^W)$. Hence, $\tilde{S}(\theta^W) - \tilde{S}(\theta^E) = \tilde{W}(\theta^W) - \tilde{W}(\theta^E) + \tilde{\Pi}(\theta^E) - \tilde{\Pi}(\theta^W) > 0$.

can continue in the case of a direct expropriation, albeit perhaps less efficiently operated by the government. Second, host countries benefit from regulatory expropriations by avoiding regulatory problems, but in the process suffer other welfare losses. With direct expropriation the benefits instead come from gaining access to a stream of operating profits. Surprisingly perhaps, these differences imply that it can actually be efficiency-enhancing to allow for direct expropriations.

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

Proposition 7 *The fully efficient solution (k^G, θ^G) 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 direct expropriations should be allowed in actual agreements, but rather that the reason for not doing so is not as trivial as it might seem. The common negative perception of direct expropriations is probably often based on the notion that they effectively constitute unproductive (or worse) thefts that deter investment. But it is exactly the fact that these measures constitute a pure transfer of ownership that might provide a role for them in investment agreements. Their role is then to mitigate overinvestment stemming from the full compensation that is used to reduce the ex post incentives to regulate.

An important caveat for efficiency enhancing direct expropriation is that the shock θ has to be observable and verifiable for a globally efficient outcome to be achievable. Assume instead that the host country has private information about θ. The value of allowing production in the foreign-owned industry is $S(k,\theta)$, whereas the value of direct expropriation is $S(k,\theta) + \Pi(k) - T^x$, where T^x is what the host country must pay in compensation to foreign investors under direct expropriation. The net benefit $\Pi(k) - T^x$ of direct expropriation is independent of the realization of θ. The host country would never truthfully reveal θ if the expropriation compensation were to depend on the shock. Hence, the only way an investment agreement can ensure 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 it impossible to have direct expropriation for some realizations of the shock and regulation for other realizations in our setting 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: $T^x = \Pi(k)$.

5.3 Monopoly

FDI is often undertaken by firms with significant market power in their output markets. Investment decisions might in such instances be influenced by how they affect the probability of regulation. Indeed, the early FDI literature discusses how investors could reduce host country governments' incentives to expropriate, for instance by choosing more complex production techniques than necessary, or by maintaining vital parts of the production process outside the host country. As pointed out above, Appendix A.2 demonstrates that Proposition 1 holds also for the case of monopoly, and with strategic investment decisions. We will now see how such market power affects the efficiency of an investment agreement.

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

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

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

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

The equilibrium threshold for regulation θ^M is thus given by $\theta^M \equiv \Theta(k^M)$.

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

Proposition 8 *Assume that the foreign investor has monopoly power, that there would be equilibrium overregulation absent any investment agreement ($\theta^M \leq \theta^G$), and that the marginal production externality is in the range*

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

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

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

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

Note that the conditions that render (θ^G, k^G) feasible under a carve-out policy are qualitatively very similar to those under perfect competition, which were established in Proposition 3: first, there is room for a carve-out policy to improve regulatory performance ($\theta^M \leq \theta^G$); second, there is under-investment because of a non-negative marginal production externality; and third, the externality is not so strong as to render full compensation optimal in all states of the world. We conclude that the properties of optimal investment agreements do not depend critically on market structure and the assumption about non-strategic investors.

5.4 Asymmetric information concerning regulatory shocks

We have thus far assumed ex post verifiability of the economic consequences of regulatory shocks. But certain aspects of investment agreements are better understood assuming that host countries have private information. For instance, Section 5.2 suggested asymmetric information as a rationale for investment agreements to have stricter compensation rules for direct than for regulatory expropriations. We will here more formally consider the design of agreements when regulatory preferences cannot be directly observed, thus allowing policy makers to misrepresent the true motives of their regulations.

With θ being observed by the host country government only, it is no longer possible to implement an agreement that awards compensation below a threshold $\hat{\theta}$, but nothing above. For realizations $\theta > \theta^N$ the host country would simply claim that $\theta > \hat{\theta}$, in order to be allowed to regulate without compensation. Incentive compatibility thus generally requires that host countries compensate firms for regulatory interventions (the proof is provided in Appendix A.8):

Proposition 9 *Assume that the host country is privately informed about the shock* θ. *For any investment agreement that satisfies Contract Restrictions (1)-(4), and for which compensation at most equals foregone operating profit* $\Pi^j(k_i)$, *there exists an alternative agreement that satisfies the same restrictions and for the host country:*
(i) features the compensation function

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

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

The optimal investment agreement under asymmetric information thus has a similar structure to the one under full information, in that it requests full compensation of operating profits $\Pi(k)$ for all shock realizations below a threshold value $\hat{\theta}$. But important differences arise for realizations $\theta > \hat{\theta}$. Under complete information, the host country will not be requested to compensate firms subsequent to regulation, whereas compensation is required in the asymmetric information case to achieve incentive compatibility. Under asymmetric information, therefore, there will be compensation payments in equilibrium, which is the case under complete information only if $\hat{\theta} > \theta^E$.

The optimal compensation scheme in Proposition 9 violates Contract Restriction (5) for $\theta > \hat{\theta}$ in two closely related respects. First, 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 for $\theta \in (\hat{\theta}, \Theta^G(k))$, since $-S(k,\hat{\theta}) < \Pi(k)$ in this case.[33] In the next Section we explore in more detail the efficiency gains that can be achieved by increasing the degree of flexibility in the compensation schemes relative to Contract Restriction (5).

5.5 Other forms of compensation schemes

The level of investment protection $\hat{\theta}$ is the only instrument under the carve-out policy (9) that can be used for correcting firms' investment incentives and the host country ex post incentive to overregulate. This single instrument is sufficient if the marginal production externality is non-negative, but not too large; see Proposition 3. However, it is necessary to introduce features that are typically not found in actual agreements to achieve full efficiency when this externality is negative, specifically by relaxing some of the constraints imposed by Contract Restrictions (1)-(5). In what follows, we first review a number of such schemes that have been analyzed in the literature, to identify how they deviate from the compensation schemes we have considered so far, and we then present an alternative efficient scheme. To facilitate comparison, we recast the other models within the context of our current framework.

Hermalin (1995) considers distortions to investments and regulation in a model with direct expropriation and a single firm. He derives two efficient mechanisms. In the first mechanism, a firm pays a production tax equal to the country's value of seizing the asset. In our setting, a tax equal to $-S(k,\theta)$ would implement the fully efficient outcome. A tax system sophisticated enough to induce each firm to internalize the full social cost of its actions would render any regulation superfluous: The firm would voluntarily shut down production whenever the social cost exceeded the benefit. The second mechanism instead requests the host country to pay the firm the same amount in compensation subsequent to expropriation. This feature highlights a fundamental property of efficient compensation, namely 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 $-\int_{\theta^G}^{\bar{\theta}} S(k,\theta)dF(\theta)$ in the

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

current setting, which generally differs from the expected compensation $-\int_{\underline{\theta}}^{\theta^G} S(k,\theta)dF(\theta)$ that generates efficient investment incentives. Efficiency requires the investor to internalize the marginal effect of the investment on the social value of allowing production rather than the value of shutting it down.

Blume et al (1984) and Aisbett et al (2010a) discuss a deviation from Contract Restriction (5), in which compensation is a linear combination of operating profits and investment costs: $T(k) = \delta\Pi(k) + \alpha R(k)$.[34] This compensation mechanism has two instruments (δ, α) that can be used to correct the distortions to investment and regulation within our framework. It is easy to verify that a special case of this compensation scheme, where

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

implements the fully efficient outcome (k^G, θ^G) as a Nash equilibrium if and only if

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

Efficiency of the carve-out policy (9) depends on verifiability of the regulatory shock θ; see the discussion in Section 5.4. A virtue of (17) is that implementation does not require that the shock θ is publicly observable.[35] However, the compensation scheme deviates from those we have previously considered in several respects. First, there are no carve-outs since compensation is paid for any regulation, in line with the scheme in Proposition 9. Second, the scheme in (17) overcompensates firms for their losses since $T(k) > \Pi(k)$ for some investments $k \neq k^G$.[36] Yet another essential feature of this scheme is that it relies on an ability to estimate the firm's operating profit $\Pi(k^G)$ and capital cost $R(k^G)$ at the efficient investment level k^G. This is of questionable empirical relevance, and it would presumably make its practical implementation difficult. Both carve-out policies and (17) are derived under the assumption that all firms are identical if an industry consists of more than one firm. If firms in an industry differ significantly in terms of size and profits, a fully efficient carve-out policy will require a unique threshold for compensation for each firm. The linear compensation rule

[34] It is of interest to examine compensation schemes based upon other factors that operating profits since the notion of "fair market value" has been interpreted in different ways in the case law. It is often interpreted to mean the value of an asset if sold to an outside party, which in principle could be based on incurred investment costs. An application of Theorem A.1 shows that linear compensation rules that incorporate both operating profit and capital costs are superior to rules that only depend on capital costs. Still, it does not seem to be a common practice to base compensation on linear combinations of foregone operating profits and incurred investment costs, which is an essential feature of the scheme here.

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

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

(17) requires firm-specific δ to ensure that each firm faces correct investment incentives. Tailoring an agreement to firm-level characteristics in this fashion can be done in commercial contracts between host countries and individual investors concerning specific projects, but does not occur in state-to-state treaties. Our analysis does not depend on firm-specific thresholds however, since we assume that all firms in an industry face the same compensation rules and are simultaneously regulated.

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

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

induces efficient investment if regulation is ex post efficient, i.e. if the host country applies the regulatory threshold $\Theta^G(k)$.[37] In particular, $T(k)$ only depends on the actual investment k. Ex post efficient regulation is ensured by requiring that the country pays compensation $\Pi(k)$. This scheme differs from those in actual agreements since it requires that the host country payment differs from the compensation received by the firm. Stähler (2016) assumes that an arbitrator enables the parties to break the payment balance in this fashion. The compensation rule is thus a Vickrey-Clarke-Groves type of mechanism.

The compensation mechanisms reviewed above all have their merits, either in terms of simplicity (carve-out policies), incentive compatibility (linear compensation as in (17)), or non-reliance on efficient investments (as in (18)). But they also have their shortcomings, either in terms of the possibility for reaching an efficient outcome or when it comes to practical implementation. We show in Appendix A.9 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; examples of such instances are the termination of the renewable energy support schemes by Spain and other countries, or the German shut down of nuclear power after Fukushima. The following result is made more precise and proved in Appendix A.9:

Proposition 10 *A compensation scheme that is based on relative performance can under certain circumstances implement full efficiency even when this cannot be done with the optimal scheme characterized in Proposition 1.*

The specific efficient compensation scheme we identify, see equation (A.18), differs from (17) by being based entirely on the investments that firms have actually made, instead of a counterfactual

[37] This is true even if the firm behaves strategically. Under the compensation rule (18), the expected profit of the firm equals $\int_{\underline{\theta}}^{\Theta^G(k)}(S(k,\theta)+\Pi(k))dF(\theta) - R(k) + \widetilde{T}$ under ex post optimal regulation, which is identical to the social welfare function up to a constant $\widetilde{T}$. The purpose of $\widetilde{T}$ is only to ensure non-negative compensation, but we do not define it here.

of what firms would have earned at the efficient outcome. Furthermore, the host country never overcompensates the firms. The rule differs from (18) 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 can implement the fully efficient outcome without information regarding the shock θ. It is robust to asymmetric information in two other dimensions as well. First, we previously assumed that policy makers in the host country ignore the effect on operating profits in the decision whether to regulate. Some of the firms could be domestically or even state-owned, or political preferences could affect the way decision makers value profit, and this could be private information (Aisbett et al, 2010a). Also, policy makers have incentives to exaggerate the extent to which they account for investor profits. Our mechanism does not depend on host country political incentives being observable; see Appendix A.9. Second, we have previously assumed that firms' operating profits are observable. But productivity differences can in practice render operating profits unobservable even if investments are the same across firms and common knowledge. Firms could then have incentives to exaggerate the value of continued production to increase their compensation. Our compensation rule is independent of the firm's own profit however, and no firm therefore has any unilateral incentive to misreport it.[38]

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

6 Concluding remarks

The number of international investment agreements has increased dramatically since the mid-1980s, and protection of foreign direct investment has become a core issue in the policy debate in developed countries. But the economic literature hardly sheds any light on their appropriate design and implications. We contribute to filling this void by examining optimal investment agreements that share basic features with actual agreements.

Our model generates a wide range of results. We see the following as particularly relevant:

- Optimal investment agreements never induce host countries to permit production when regulation would increase joint welfare; there is no *global regulatory chill*;

[38] Myerson and Satterthwaite (1983) derive an optimal compensation mechanism with asymmetric information on both sides when there is a single firm. Their compensation scheme features payments even if there is no regulation and therefore violates Contract Restriction (4). They also consider the case where an arbitrator breaks the payment balance.

- A simple compensation scheme based solely on foregone operating profits can fully resolve the distortions in investment and regulation under a robust set of circumstances;
- The incentives to form investment agreements depend on the ability of host countries to make unilateral commitments with regard to investment protection, as well as on the direction of the potential investment flows between the countries.
- Investment agreements between developed economies, such as TTIP, are likely to have strong distributional implications by benefiting foreign investors at the expense of rest of society. Agreements between developed and developing countries do not have these distributional effects, but are formed only if they increase domestic welfare in the developing countries.

We conclude by suggesting some avenues for future research. A common critique in the policy debate concerns the legal standing of foreign investors in investment agreements through the ISDS mechanisms. A key issue here is how a system that only allows foreign governments to litigate would differ from current ISDS systems. It seems intuitively plausible that ISDS systems imply more active enforcement, but this is only a conjecture.

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. However, the precise form of interaction is unclear; see Maggi (2016) for an analytical taxonomy of various forms of complementarities between undertakings in trade agreements. Our paper considers investment agreements in isolation.

Extending the model to include a process of lobbying, as in Grossman and Helpman (1994, 1995), could further illuminate the politics of investment agreements. We simply assume that an agreement is formed if and only if it increases welfare in each country.

In our model, investors' only choice is whether or not to make a direct investment. An alternative could be to establish an arms-length arrangement with a local producer. Arms-length contracts typically are incomplete and sometimes associated with hold-up problems between firms, as highlighted in the literature on outsourcing (Helpman, 2006, and Antràs and Rossi-Hansberg, 2009). An advantage of arms-length contracts could be that a local producer is less likely to be regulated than a vertically integrated foreign firm. Investment agreements could thus make outsourcing less attractive through increased protection of foreign investment.

We consider the incentives to form a single bilateral investment agreement. But it is sometimes maintained that the surge of bilateral investment treaties between developed and developing countries constitutes a race between developing countries to attract foreign investment. It therefore seems relevant to extend the analysis to a setting with competition for investment between countries.

The analysis focuses for the most part on agreements fulfilling Contract Restrictions (1)-(5). There is strong institutional support for these assumptions. But it might nevertheless be interesting to endogenize some of these incomplete contracting features, possibly within a contracting cost framework similar to the one developed by Horn et al (2010) to endogenize the design of trade agreements.

Finally, it is natural to start the analysis of investment agreements by analyzing the design and implications of voluntary, optimal agreements. But agreements might in practice contain provisions that are harmful to some of the involved partied because of poor drafting of the agreements, or because of undesirable interpretations by arbitration panels. Much of the very recent literature on investment agreements examines the consequences of exogenously imposed provisions. We view these approaches as complementary to the present one.

References

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

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

Antràs, Pol and Esteban Rossi-Hansberg (2009). "Organizations and Trade." *Annual Review of Economics 1,* 43-64.

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), February, 71-92.

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

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

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

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

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

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

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

Grossman, Gene M. and Elhanan Helpman (1994). "Protection for Sale. *The American Economic Review* 84(4), 833-850.

Grossman, Gene M. and Elhanan Helpman (1995). "Trade Wars and Trade Talks." *Journal of Political Economy* 103(4), 675-708.

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

Helpman, Elhanan (2006). "Trade, FDI, and the Organization of Firms." *Journal of Economic Literature* 44(3), 589-630.

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.

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." Mimeo, April 22.

Konrad, Kai A. (2016). "Large Investors and Permissive Regulation: Why Environmentalists May Dislike Investor-State Dispute Settlement." Max Planck Institute for Tax Law and Public Finance Working Paper 2016-10.

Maggi, Giovanni (2016). "Issue Linkage." In K. Bagwell and R.W. Staiger (eds.), *The Handbook of Commercial Policy*, Elsevier.

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.

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

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

Schnitzer, Monica (1999). "Expropriation and Control Rights: A Dynamic Model of Foreign Direct Investment." *International Journal of Industrial Organization* 17(8), 1113-1137.

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.

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

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

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

UNCTAD (2015). *World Investment Report.*

Vernon, Raymond (1971). *Sovereignty at Bay: The Multinational Spread of US Enterprises.* New York: Basic Books.

A Appendix

A.1 Properties of $X^i(k_i)$, $P^i(k_i)$ and $K^i(\hat{\theta}_i)$

The expressions in (1) yield

$$X_k^i(k_i) = \frac{-C_{xk}^j(X^i(k_i), k_i)}{C_{xx}^j(X_k^i(k_i), k_i) - \Omega_{zz}^i(X_k^i(k_i))} > 0,\ P_k^i(k_i) = \Omega_{zz}^i(X^i(k_i))X_k^i(k_i) < 0.$$

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

$$K_\theta^i = \frac{-f^i(\hat{\theta}_i)C_k^j(X^i(\hat{k}_i), \hat{k}_i)}{F^i(\hat{\theta}_i)[C_{xk}^j(X^i(\hat{k}_i), \hat{k}_i)X_k^i(\hat{k}_i) + C_{kk}^j(X^i(\hat{k}_i), \hat{k}_i)] + R_{kk}^j(\hat{k}_i)} > 0.$$

Monotonicity of $K^i(\hat{\theta}_i)$ follows from $C_k^j < 0$, the concavity of consumer utility and convexity of the production and investment cost:

$$C_{xk}^j X_k^i + C_{kk}^j = \frac{C_{xx}^j C_{kk}^j - C_{xk}^j C_{kx}^j - \Omega_{zz}^i C_{kk}^j}{C_{xx}^j - \Omega_{zz}^i} > 0.$$

A.2 Properties of optimal compensation schemes

This Appendix proves statements made in Section 4 concerning properties of investment agreements, using more general frameworks than those employed in the main text in several respects. First, it makes more general assumptions concerning industry structure, for instance by allowing firms to be heterogenous in various ways, and to be imperfectly competitive. Second, it considers a range of different scenarios with regard to contract restrictions, and to the behavior of investors. The Appendix establishes that in each of these scenarios, any optimal investment agreement can be characterized in terms of a threshold that yields domestic, but never global regulatory chill. For certain scenarios we also establish properties of optimal compensation schemes.

Sections A.2.2 and A.2.3 focus on compensation schemes fulfilling Contract Restrictions (1)-(4), but not necessarily Restriction (5). Proposition A.1 in Section A.2.2 assumes that firms invest strategically to influence ex post regulatory decisions. A generalization of Proposition 1 in the main text is provided in Proposition A.2 in Section A.2.3, which considers the case where firms do not behave strategically. We then consider the consequences of more restrictive compensation mechanisms that limit compensation to be proportional and at most equal to operating profit. Proposition A.3 in Section A.2.4 shows that the optimal compensation function is a carve-out policy if investors behave strategically, and Proposition A.4 in Section A.2.5 establishes the same result for non-strategic investors. Proposition A.5 in Section A.2.6 verifies the optimality of a carve-out policy under a monopoly market structure under the restriction that compensation is limited to at most operating profit, but compensation is not required to be proportional.

A.2.1 A generalized model

There are two countries, indexed by $i \neq j = 1, 2$, and an industry with $H \geq 1$ firms, indexed by $h = 1, 2..., H$. 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. We assume that firms make their investment decisions simultaneously and independently to maximize unilateral profit, but do not make any assumptions about the nature of strategic interaction in the investment stage nor in the product market. Let $\Pi^{hi}(\mathbf{k}_i) \geq 0$ be the reduced form operating profit of firm h of its facilities in country i, and assume that this profit 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 either $k_{h1} > 0$ or $k_{h2} > 0$.

Let the reduced form domestic welfare be $S^i(\mathbf{k}_i, \theta_i)$ under production and zero if there is regulation. This domestic welfare depends on domestic investment $\mathbf{k}_i$ and on 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 means that investments are strictly positive for all firms). Assume that $S^i(\mathbf{k}_i, \theta_i)$ and the profit functions are continuous in $\mathbf{k}_i$.

Assume that θ_i is continuously distributed on $[\underline{\theta}_i, \bar{\theta}_i]$ with marginal cumulative distribution function $F^i(\theta_i)$ and marginal 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 the production of all firms in the industry in the host country. Both countries take the decision to regulate simultaneously and independently.

Consider now the ex post optimal choice of country i whether to allow production absent an investment agreement. 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

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

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 the implicit solution to $\Delta^i(\mathbf{k}_i, \Theta^i, \boldsymbol{\gamma}_i) \equiv 0$ in the intermediate case. We assume that the host country allows production if indifferent. Country i will then allow production if and only if $\theta_i \leq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. The decision to regulate 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 $\Theta^{iG}(\mathbf{k}_i) \equiv \Theta^i(\mathbf{k}_i, \mathbf{1}) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. This is the cut-off below which it is ex post optimal to allow production in country i if national welfare is defined by the sum of domestic welfare and industry operating profit.

An *international investment agreement* (IIA) is a vector $\hat{\mathbf{T}}^i = (\hat{T}^{1i}, .., \hat{T}^{hi}, .., \hat{T}^{Hi})$ of compensation rules for each country, where $\hat{T}^{hi}(\mathbf{k}_i, \theta_i) \geq 0$ 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; a country thus never compensates for regulation abroad. .

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

A.2.2 General compensation schemes: Strategic investors

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

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

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

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

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

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

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

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

i equals

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

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

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

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

$$\tilde{V}^i(\hat{\mathbf{T}},\boldsymbol{\gamma}_i)=\tilde{W}^i(\hat{\mathbf{T}})+\sum_{h=1}^H[\gamma_{hi}(\tilde{\Pi}^h(\hat{\mathbf{T}})-R^h(\hat{\mathbf{k}}_h))-\tilde{\Pi}^{hi}(\hat{\mathbf{T}})]. \tag{A.5}$$

Proposition A.1 *Assume that all firms account for the effect of their investment on regulation. For any investment agreement that satisfies Contract Restrictions (1)-(4) there exists an alternative agreement that satisfies the same restrictions, and that for each country i:*
(i) implements a threshold function for regulation $\Theta^{i*}(\mathbf{k}_i,\boldsymbol{\gamma}_i)\in[\Theta^i(\mathbf{k}_i,\boldsymbol{\gamma}_i),\Theta^{iG}(\mathbf{k}_i)]$*;*
(ii) yields weakly higher expected welfare and industry profits than the initial agreement.

Proof: The method of proof is to show that for any IIA with compensation rule $\hat{\mathbf{T}}$ satisfying the appropriate restrictions, there exists another IIA with compensation rule $\mathbf{T}$ satisfying the same restrictions, with the characteristics in the theorem and that yields weakly higher expected domestic welfare and industry profits than the initial agreement.

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

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

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

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

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

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

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

and where

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

if $\int_{B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}_i) > 0$. This alternative agreement builds on the payments under the original compensation scheme and the operating profits of regulated firms. Θ^{i*} therefore depends on $\boldsymbol{\gamma}_i$ since $M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ depends on $\boldsymbol{\gamma}_i$.

Establishing $\Theta^{i*}(\mathbf{k}, \boldsymbol{\gamma}_i) \in [\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i), \Theta^{iG}(\mathbf{k}_i)]$. The inequality $\Theta^{i*}(\mathbf{k}, \boldsymbol{\gamma}_i) \leq \Theta^{iG}(\mathbf{k}_i)$ follows directly from (A.6). Furthermore, $\Theta^{i*}(\mathbf{k}, \boldsymbol{\gamma}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ trivially holds if $\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) = \underline{\theta}_i$. To establish $\Theta^{i*}(\mathbf{k}, \boldsymbol{\gamma}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ for $\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) > \underline{\theta}_i$, note that if $F^i(\Theta^{iG}(\mathbf{k}_i)) \leq \int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i)$, then $F^i(\Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) = F^i(\Theta^{iG}(\mathbf{k}_i)) \geq F^i(\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i))$, where the inequality follows from $\Theta^{iG}(\mathbf{k}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. Assume finally that $\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i) \leq F^i(\Theta^{iG}(\mathbf{k}_i)$. The assumption that $S^i(\mathbf{k}_i, \theta_i)$ is strictly decreasing in θ_i and $\hat{T}^{hi}(\mathbf{k}_i, \theta_i) \geq 0$ jointly imply that

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

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

Country i allows production under agreement T iff $\theta_i \leq \Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. Consider the incentives for the host country to regulate the industry under an arbitrary investment profile $\mathbf{k}_i$ for agreement $\mathbf{T}$ and for different realizations of the shock θ_i:
(i) $\theta_i \in A^i(\mathbf{k}_i, \hat{\mathbf{T}}^i) \cup A^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i) = [\underline{\theta}_i, \Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i)]$. By construction of the agreement, the net benefit of allowing production is non-negative for all $\theta_i \leq \Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i) \leq \Theta^{iG}(\mathbf{k}_i)$ because in this case

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

(ii) $\theta_i \in B^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)$. It is optimal to regulate because the compensation function remains the same as before, and it was optimal to regulate already under the initial agreement.
(iii) $\theta_i \in B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)$ and $\int_{B^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\tilde{\theta}^i) > 0$. By the construction of $\Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$:

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

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

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

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

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

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

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

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

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

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

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

under agreement $\mathbf{T}$, where

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

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

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

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

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

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

We have thus shown that for any IIA with arbitrary non-negative compensation $\hat{\mathbf{T}}$ that is paid if and only if the host country disallows production, we can find another compensation rule $\mathbf{T}$ that is paid if and only if the host country disallows production, that increases regulatory efficiency, but without affecting equilibrium investments. We characterized one such compensation rule in (A.7)-(A.8), but many other compensation rules can sustain the same result.

Our specific compensation rule yields a compensation T^{hi} to firm h in country i that is a convex combination of that firm's operating profit Π^{hi} and the compensation $\hat{T}^{hi}$ in the original scheme, where the weights on the two components are country-specific and depend on θ_i, but are the same for all firms that have invested in country i. This structure implies that the modified scheme $\mathbf{T}$ inherits a number of characteristics from the original scheme $\hat{\mathbf{T}}$. First, compensation is non-negative because operating profit is non-negative and the original compensation is non-negative ($\Pi^{hi} \geq 0$ and $\hat{T}^{hi} \geq 0$ imply $T^{hi} \geq 0$). Second, it does not rely on excessive compensation (punitive damages) if this is not part of the original scheme ($\hat{T}^{hi} \leq \Pi^{hi}$ implies $T^{hi} \leq \Pi^{hi}$). 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 those imposed by Contract Restrictions (1)-(4), and the non-negative compensation requirement. 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 5.5.

A.2.3 General compensation schemes: Non-strategic investors

The above results are based on the assumption that firms take into account how their investments affect the probability of being regulated. In this 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 (NE) is the appropriate equilibrium concept. Given the investment agreement $\hat{\mathbf{T}}$, an NE defines two subsets of shock realizations in each country, the set $\hat{M}^i$ of θ_i for which the host country allows production and the complementary set $\hat{M}^{ir}$ of θ_i for which the host country 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:

$$\begin{aligned} \hat{M}^i \equiv & \quad \{\theta_i : \Delta^i(\hat{\mathbf{k}}_i, \theta_i, \boldsymbol{\gamma}_i) + \textstyle\sum_{h=1}^{H}(1-\gamma_{hi})\hat{T}^{hi}(\hat{\mathbf{k}}_i, \theta_i)\} \geq 0\}, \\ \hat{M}^{ir} \equiv & \quad \{\theta_i \notin \hat{M}^i, \end{aligned} \tag{A.10}$$

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 $\hat{M}^{ir}$:

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

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

Proposition A.2 *Assume that all firms treat regulation as exogenous to their own investment. For any investment agreement that satisfies Contract Restrictions (1)-(4) there exists an alternative agreement that satisfies the same restrictions, and that for each country i:*
(i) implements a threshold function for regulation $\Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i) \in [\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i), \Theta^{iG}(\mathbf{k}_i)]$*;*
(ii) yields weakly higher expected welfare and industry profits than the initial agreement.

Proof: For any initial agreement $\hat{\mathbf{T}}$, define the modified agreement $\mathbf{T}$ by (A.6)-(A.8). It is then optimal for country i to allow production if and only if $\theta_i \leq \Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i) \in [\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i), \Theta^{iG}(\mathbf{k}_i)]$ 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

$$F^i(\hat{\theta}_i^*) \equiv \min\{\textstyle\int_{\hat{M}^i} dF^i(\theta_i); F^i(\hat{\theta}_i^G)\}.$$

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

$$\begin{aligned} &\Pi^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}) F^i(\hat{\theta}_i^*) + \tilde{T}^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}) \textstyle\int_{\hat{B}^i} dF^i(\theta_i) + \int_{\hat{B}^{ir}} \hat{T}^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}, \theta_i) dF^i(\theta_i) \\ &= \Pi^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}) \textstyle\int_{\hat{M}^i} dF^i(\theta_i) + \int_{\hat{M}^{ir}} \hat{T}^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi}, \theta_i) dF^i(\theta_i) \end{aligned}$$

under the modified agreement $\mathbf{T}$, given the investment profile $\hat{\mathbf{k}}_{-hi}$ of all other firms active in country i and the expectation that $A^i(k_{hi}, \hat{\mathbf{k}}_{-hi}, \hat{\mathbf{T}}^i) = \hat{A}^i$, $A^{ir}(k_{hi}, \hat{\mathbf{k}}_{-hi}, \hat{\mathbf{T}}^i) = \hat{A}^{ir}$, 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 $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$. And since the domestic welfare, all operating profits and investment are the same as before, it follows that the alternative agreement $\mathbf{T}$ represents an expected improvement for all parties even under Nash implementation.■

Proposition 1 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 theorems are silent about the optimal compensation for $\theta_i > \hat{\Theta}^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ because the modified compensation scheme $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$ is defined relative to some initial and arbitrary compensation scheme $\hat{\mathbf{T}} = (\hat{\mathbf{T}}^1, \hat{\mathbf{T}}^2)$ in this case. To obtain sharper results in this regard we need to place more structure on permissible compensation schemes. To this end, we require that agreements fulfil Contract Restriction (5), stipulating *proportional* compensation schemes:

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

A.2.4 Proportional compensation schemes: Strategic investors

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

$$\begin{aligned}\hat{\mathbf{k}}_h \in\ & \underset{\mathbf{k}_h \in \mathbb{R}_+^2}{\arg\max}\{\Pi^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi})[\textstyle\int_{M^i(k_{hi},\hat{\mathbf{k}}_{-hi},\hat{\mathbf{T}}^i)} dF^i(\theta_i) + \int_{M^{ir}(k_{hi},\hat{\mathbf{k}}_{-hi},\hat{\mathbf{T}}^i)} b_i(\theta_i) dF^i(\theta_i)] \\ +\ & \Pi^{hj}(k_{hj}, \hat{\mathbf{k}}_{-hj}) \textstyle\int_{M^j(k_{hj},\hat{\mathbf{k}}_{-hj},\hat{\mathbf{T}}^j)} dF^j(\theta_j) + \int_{M^{jr}(k_{hj},\hat{\mathbf{k}}_{-hj},\hat{\mathbf{T}}^i)} \hat{T}^{hj}(k_{hj}, \hat{\mathbf{k}}_{-hj}, \theta_j) dF^j(\theta_i) - R^h(\mathbf{k}_h)\}\end{aligned}$$

for all $h = 1, 2..., H$. In this case, Proposition A.1 can be tightened considerably:

Proposition A.3 *Assume that all firms account for the effect of their investment on regulation. For any investment agreement that satisfies Contract Restrictions (1)-(5) there exists an alternative agreement that satisfies the same restrictions, and that for each country i:*
(i) features the following compensation function for each firm h

$$T^{hi}(\mathbf{k}_i, \theta_i) \equiv \begin{cases} \Pi^{hi}(\mathbf{k}_i) & \theta_i \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \\ 0 & \theta_i > \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i). \end{cases} \tag{A.13}$$

(ii) implements a threshold function for regulation $\Theta^{i*}(k_i) \in \{\Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i); \Theta^{iG}(\mathbf{k}_i)\}$*;*
(iii) yields weakly higher expected domestic welfare and industry profits than the initial agreement.

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

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

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

$$F^i(\Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) = \textstyle\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta) \leq F^i(\Theta^{iG}(\mathbf{k}_i)). \tag{A.14}$$

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

$$F^i(\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)) \equiv \textstyle\int_{M^i(\mathbf{k}_i, \hat{\mathbf{T}}^i)} dF^i(\theta_i) + \textstyle\int_{M^{ir}(\mathbf{k}_i, \hat{\mathbf{T}}^i)} b(\theta_i) dF^i(\theta_i) \leq 1 \tag{A.15}$$

and the compensation mechanism (A.13).

Establishing $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. A comparison of (A.15) and (A.14) yields $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \geq \Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, whereas $\Theta^{i*}(\mathbf{k}_i, \boldsymbol{\gamma}_i)) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ by the assumption that compensation is non-negative; see the proof of Proposition A.1.

Country i allows production under $\mathbf{T}^i$ iff $\theta_i \leq \min\{\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i); \Theta^{iG}(\mathbf{k}_i)\}$. The net benefit of allowing production given $\mathbf{T}$ equals

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

if $\theta_i \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ and is non-negative if $\theta_i \leq \Theta^{iG}(\mathbf{k}_i)$. We have already shown that it is ex post optimal for the host country to disallow production for all $\theta_i > \Theta^{iG}(\mathbf{k}_i)$ if compensation to each firm is at most $\Pi^{hi}(\mathbf{k}_i)$. If $\Theta^{iG}(\mathbf{k}_i) \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, then we are done. If $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) < \Theta^{iG}(\mathbf{k}_i)$, then the net benefit of allowing production is given by $\Delta^i(\mathbf{k}_i, \theta_i, \boldsymbol{\gamma}_i)$ in the range $\theta_i \in (\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i), \Theta^{iG}(\mathbf{k}_i)]$, which is strictly negative by $\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i) \geq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. We conclude that it is ex post optimal for country i to allow production if and only if $\theta_i \leq \min\{\hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i); \Theta^{iG}(\mathbf{k}_i)\}$ if all firms receive compensation according to T^{hi} defined in (A.13).

Investments and profits are the same under the two agreements. Firm h receives its operating profit $\Pi^{hi}(\mathbf{k}_i)$ for all $\theta_i \leq \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ independently of whether it is regulated or not. It is regulated, but receives no compensation for all $\theta_i > \hat{\Theta}^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$. Hence, firm h's expected operating profit in country i equals

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

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

Equilibrium operating profits and investments are independent of whether country i offers $\hat{\mathbf{T}}^i$ or $\mathbf{T}^i$, so $\tilde{\Pi}^{hi}(\mathbf{T}^i, \hat{\mathbf{T}}^j) = \tilde{\Pi}^{hi}(\hat{\mathbf{T}})$, $\tilde{\Pi}^{hj}(\hat{\mathbf{T}}^j, \mathbf{T}^i) = \tilde{\Pi}^{hj}(\hat{\mathbf{T}})$ and $\tilde{\Pi}^{h}(\mathbf{T}^i, \hat{\mathbf{T}}^j) = \tilde{\Pi}^{h}(\hat{\mathbf{T}})$ for all h.

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

$$\tilde{V}^i(\mathbf{T}^i, \hat{\mathbf{T}}^j, \boldsymbol{\gamma}_i) \equiv \tilde{W}^i(\mathbf{T}^i, \hat{\mathbf{T}}^j) + \textstyle\sum_{h=1}^{H}(\gamma_{hi}(\Pi^h(\hat{\mathbf{T}}) - R^h(\hat{\mathbf{k}}_h)) - \Pi^{hi}(\hat{\mathbf{T}}))$$

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

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

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

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

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

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

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

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

by (A.15) and

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

by (A.14). Hence, $V^i(\mathbf{T}^i, \hat{\mathbf{T}}^j, \boldsymbol{\gamma}_i) \geq V^i(\hat{\mathbf{T}}, \boldsymbol{\gamma}_i)$. It follows that a unilateral deviation by country i from $\hat{\mathbf{T}}^i$ to $\mathbf{T}^i$ represents an improvement in the expected welfare of country i without affecting the welfare in country j or the expected profits of firms negatively. Analogously, we can show the same result for a unilateral deviation by country j from $\hat{\mathbf{T}}^j$ to $\mathbf{T}^j$ if even country j is restricted to proportional compensation mechanisms. Hence, $\mathbf{T} = (\mathbf{T}^1, \mathbf{T}^2)$ is a (weakly) better policy than $\hat{\mathbf{T}}$ for both parties to an international investment agreement and to firms under the restriction to proportional compensation.■

A.2.5 Proportional compensation schemes: Non-strategic investors

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

$$\begin{aligned}\hat{\mathbf{k}}_h \in \; & \underset{\mathbf{k}_h \in \mathbb{R}^2_+}{\arg\max} \{\Pi^{hi}(k_{hi}, \hat{\mathbf{k}}_{-hi})[\textstyle\int_{\hat{M}^i} dF^i(\theta_i) + \int_{\hat{M}^{ir}} b_i(\theta_i) dF^i(\theta_i)] \\ + \; & \Pi^{hj}(k_{hj}, \hat{\mathbf{k}}_{-hj}) \textstyle\int_{\hat{M}^j} dF^j(\theta_j) + \int_{\hat{M}^{jr}} \hat{T}^{hj}(k_{hj}, \hat{\mathbf{k}}_{-hj}, \theta_j) dF^j(\theta_i) - R^h(\mathbf{k}_h)\}.\end{aligned}$$

Proposition A.4 *Assume that all firms treat regulation as exogenous to their own investment. For any investment agreement that satisfies Contract Restrictions (1)-(5) there exists an alternative agreement that satisfies the same restrictions, and that for each country i:*
(i) features the compensation function

$$T^{hi}(\mathbf{k}_i, \theta_i) = \begin{cases} \Pi^{hi}(\mathbf{k}_i) & \text{if } \theta_i \leq \hat{\theta}_i \\ 0 & \text{if } \theta_i > \hat{\theta}_i. \end{cases} \tag{A.16}$$

(ii) implements a threshold function for regulation

$$\Theta^{i*}(k_i) \equiv \begin{cases} \Theta^i(k_i) & \text{if } \hat{\theta}_i < \Theta^i(k_i) \\ \hat{\theta}_i & \text{if } \hat{\theta}_i \in [\Theta^i(k_i), \Theta^{iG}(k_i)] \\ \Theta^{iG}(k_i) & \text{if } \hat{\theta}_i > \Theta^{iG}(k_i); \end{cases}$$

(iii) yields weakly higher expected domestic welfare and foreign industry profits than the initial agreement.

Proof: For an arbitrary compensation rule $\hat{\mathbf{T}}^i$, consider the properties of an alternative compensation rule $\mathbf{T}^i$ in country i characterized in terms of a threshold $\hat{\theta}_i$ given by

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

and where the compensation to each firm is characterized by (A.16).

We already know from the proof of Proposition A.1 that it is optimal for country i to allow production for all $\theta_i \leq \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$ for any mechanism with non-negative compensation. In the proof of Proposition A.3, we also showed that it is ex post optimal to regulate for all $\theta_i > \hat{\Theta}^{iG}(\mathbf{k}_i)$ for any mechanism that restricts the host payment to at most the industry operating profit. If $\hat{\theta}_i > \Theta^i(\mathbf{k}_i, \boldsymbol{\gamma}_i)$, then it is optimal to allow production for $\theta_i \leq \min\{\hat{\theta}_i; \hat{\Theta}^{iG}(\mathbf{k}_i)\}$ under the proportional mechanism because then

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

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

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

With the anticipated regulation level $\hat{\theta}_i$ and proportional compensation rule $T^{hi}(\mathbf{k}_i, \theta_i)$ in country i, and given $\hat{\mathbf{T}}^j$ in country j, it is still optimal for firm h to invest $\hat{\mathbf{k}}_h$ if the other firms maintain their investments at the same level as before. And since the equilibrium investments, the expected operating profits and the domestic welfare in country i are the same as in Proposition A.3, the implications for expected country welfare and operating profits follow.■

Proposition 2 in the main text is a special case of Proposition A.4 above, with $b_i(\theta_i) \in \{0, 1\}$, one representative firm in each country investing only in FDI, and where $\gamma_{hi} = 0$ for the foreign firm h investing in country i.

A.2.6 Restricted compensation schemes: Monopoly

Reduce the number of firms in the industry in country i to one, and assume that this monopoly accounts for the effect on regulation when it decides how much to invest. Now a carve-out policy is optimal within a broader class of rules than those that require proportional compensation (A.12):

Proposition A.5 *Assume that the industry in country i consists of a monopoly that behaves strategically. For any investment agreement that satisfies Contract Restrictions (1)-(4), and for which compensation at most equals foregone operating profit $\Pi^j(k_i)$, there exists an alternative agreement that satisfies the same five restrictions, and that for each country i:*
(i) features the compensation function

$$T^i(k_i, \theta_i) \equiv \begin{cases} \Pi^j(k_i) & \theta_i \leq \hat{\Theta}^i(k_i, \gamma_i) \\ 0 & \theta_i > \hat{\Theta}^i(k_i, \gamma_i). \end{cases} \tag{A.17}$$

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

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

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

and the compensation mechanism by (A.17). The proofs that $\hat{\Theta}^i(k_i, \gamma_i) \geq \Theta^i(k_i, \gamma_i)$ and that country i allows production if and only if $\theta_i \leq \min\{\hat{\Theta}^i(k_i, \gamma_i); \Theta^{iG}(k_i)\}$ are analogous to the proofs of the same results in Proposition A.3 and therefore omitted.

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

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

which is the same expected operating profit in country i as under the original agreement $\hat{T}^i$ for every possible investment profile k_i. Neither $\hat{\mathbf{T}}^j$ 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 $(T^i, \hat{\mathbf{T}}^j)$.

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 T^i than $\hat{T}^i$ are identical to those in Proposition A.3. Hence, it is unnecessary to repeat those steps here.■

A.3 Proof of Lemma 1

Let $\hat{\theta}_i^{\mathcal{B}}$ be the firm's consistent belief about investment protection and $\hat{k}_i^{\mathcal{B}} \equiv K^i(\hat{\theta}_i^{\mathcal{B}})$ its profit maximizing investment subsequent to the announcement of $\hat{\theta}_i \leq \theta_i^N$. The firm will earn its full operating profit if $\theta_i \leq \max\{\hat{\theta}_i; \Theta^i(\hat{k}_i^{\mathcal{B}})\}$ and obtain zero profit otherwise. Hence, the firm's beliefs about investment protection is consistent with host country regulation only if $\hat{\theta}_i^{\mathcal{B}} \in \{\hat{\theta}_i; \theta_i^N\}$ because $\hat{\theta}_i^{\mathcal{B}} = \Theta^i(\hat{k}_i^{\mathcal{B}})$ if and only if $\hat{\theta}_i^{\mathcal{B}} = \theta_i^N$ by assumption (7). Assume that $\hat{\theta}_i < \theta_i^N$ and suppose $\hat{\theta}_i^{\mathcal{B}} = \hat{\theta}_i$. In this case, the host country optimally permits production if and only if $\theta_i \leq \Theta^i(K^i(\hat{\theta}_i)) > \hat{\theta}_i = \hat{\theta}_i^{\mathcal{B}}$, which is inconsistent. Hence, the only candidate for consistent beliefs is $\hat{\theta}_i^{\mathcal{B}} = \theta_i^N$ for $\hat{\theta}_i \leq \theta_i^N$. The optimal investment then equals $k_i^N = K^i(\theta_i^N)$, and the threshold for regulation occurs at $\Theta^i(K^i(\theta_i^N)) = \theta_i^N$, which verifies consistency in this final case.■

A.4 Proof of Lemma 2

Consider first the properties of θ_{NTOnly}^{NN}. Observe that

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

implies a welfare difference

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

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

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

which implies a welfare difference

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

which is strictly positive for all $\hat{\theta} > \theta^W$. Hence, $\theta^{NN}_{NTOnly} \leq \theta^W$. To establish strict inequalities, assume that $\theta^{NN}_{NTOnly} \in (\theta^N, \bar{\theta})$. It is obviously the case that $\theta^{NN}_{NTOnly} > \theta^U$ if $\theta^U = \theta^N$, but $\theta^{NN}_{NTOnly} > \theta^U$ also if $\theta^U > \theta^N$ because then

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

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

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

Consider next the properties of θ^{NS}_{NT}:

$$\tilde{S}(\theta^{NN}_{NTOnly}) + \tilde{W}(\theta^{NN}_{NTOnly}) \geq \tilde{S}(\hat{\theta}) + \tilde{W}(\hat{\theta})$$

and $\tilde{\Pi}(\theta^{NN}_{NTOnly}) > \tilde{\Pi}(\hat{\theta})$ for all $\hat{\theta} < \theta^{NN}_{NTOnly}$ imply $\theta^{NS}_{NT} \geq \theta^{NN}_{NTOnly}$. The inequality is strict if $\theta^{NN}_{NTOnly} \in (\theta^N, \bar{\theta})$ because then

$$\tilde{S}_\theta(\theta^{NN}_{NTOnly}) + \tilde{W}_\theta(\theta^{NN}_{NTOnly}) = 0,$$

but $\tilde{\Pi}_\theta(\theta^{NN}_{NTOnly}) > 0$.■

A.5 Proof of Proposition 3

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

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

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

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

Hence, $\tilde{W}(\hat{\theta})$ is strictly quasi-concave in the domain $\hat{\theta} \geq \theta^E$.

Part (a): The marginal expected welfare is strictly negative for all $\hat{\theta} \geq \theta^E$ if (13) is satisfied: $\tilde{W}_\theta(\hat{\theta}) \leq \tilde{W}_\theta(\theta^E) < 0$. Hence, $\theta^W < \theta^E$. We already know from Lemma 1 that $\theta^W > \theta^N$. By the stability condition (7), it follows that $\theta^W \in (\Theta(k^W), \Theta^G(k^W))$.

Part (b): The marginal expected welfare satisfies $\tilde{W}_\theta(\theta^E) \geq 0$ and $\tilde{W}_\theta(\bar{\theta}) \leq 0$ if (13) is violated, but (14) is satisfied. In this case, there exists a $\theta^W \in [\theta^E, \bar{\theta}]$ such that $\tilde{W}_\theta(\theta^W) = 0$. As k^G is the unique welfare maximizing investment when regulation is ex post efficient, it follows that $k^W = K(\theta^W) = k^G$. Furthermore, $\theta^W \geq \Theta^G(k^W) > \Theta(k^W)$ by stability (7) implies that the host country threshold for regulation is $\min\{\theta^W; \Theta^G(k^W)\} = \Theta^G(k^W) = \Theta^G(k^G) = \theta^G$.

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

A.6 Proof of Proposition 7

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

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

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

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^G fulfills the first–order condition (3).■

A.7 Proof of Proposition 8

Let κ be the set of k satisfying $\hat{\theta}^M \geq \Theta(k)$. Observe that $k^G \in \kappa$ by assumption because $\Theta(k^G) \leq \theta^G \leq \hat{\theta}^M$. It is ex post optimal for the host country to allow production for investments $k \in \kappa$ if and only if $\theta \leq \min\{\hat{\theta}^M; \Theta^G(k)\}$. Hence, the expected monopoly profit equals $F(\hat{\theta}^M)\Pi(k) - R(k) = \frac{R_k(k^G)}{\Pi_k(k^G)}\Pi(k) - R(k)$ in the domain κ. k^G is the profit-maximizing investment in κ because the expected monopoly profit is strictly concave in this domain, and k^G is the unique solution to the first-order condition $\frac{R_k(k^G)}{\Pi_k(k^G)}\Pi_k(k) - R_k(k) = 0$. Let $\pi^G \equiv \frac{R_k(k^G)}{\Pi_k(k^G)}\Pi(k^G) - R(k^G)$. For investments $k \notin \kappa$, it is ex post optimal for the host country to allow production if and only if $\theta \leq \Theta(k)$. Hence, the expected monopoly profit equals $F(\Theta(k))\Pi(k) - R(k) \leq F(\theta^M)\Pi(k^M) - R(k^M) \equiv \pi^M$ for all $k \notin \kappa$. It follows that k^G maximizes the expected profit for all $k \geq 0$ because

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

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

A.8 Proof of Proposition 9

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

The equilibrium rent of the host country is

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

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

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

The incentive compatible compensation is therefore given by

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

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

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

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

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

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

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

which we have ruled out by assumption.

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

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

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

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

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

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

depending on its report.

To see that this compensation scheme yields the same outcome as above, assume first that $\hat{\theta} \leq \Theta(k)$. The country would always report $\theta > \hat{\theta}$ subsequent to regulation in order to pay zero compensation: $\max\{-S(k,\hat{\theta});0\} = 0$. 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 $\theta > \Theta(k)$.

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

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

A.9 A compensation scheme based 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 can then write demand, price, consumer surplus and so forth as functions of $\mathbf{k}$. In particular, the operating profit of firm h is $\Pi^h(\mathbf{k}) \equiv \hat{\Pi}(P(\mathbf{k}), k_h)$. We maintain the assumption that firms are price-takers, so that $-C_k(X(k_h, \mathbf{k}_{-h}), k_h)$ represents the marginal perceived effect of increasing investment k_h on firm h's operating profit, where $X(k_h, \mathbf{k}_{-h})$ is the production of firm 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.

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

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

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

Let

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

be the expected externality 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' externality as exogenous to the own investment k_h.[39]

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

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

yields non-negative compensation for all firms in the industry. Here, the compensation depends not only on operating profit, but also on the externality. For instance, the firm receives a relatively large compensation if the externality of its investment is positive compared to that of the other firms in the industry.[40]

[39] Independence is a behavioral assumption here, but could be affected by technology. If the externality is additive, $\Psi(\mathbf{k},\theta) \equiv \sum_{h=1}^{I} \hat{\Psi}^h(k_h,\theta)$, 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.

[40] The compensation rule (A.18) 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 $|\mathcal{H}(\mathbf{k})| = 1$.

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 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.

The expected profit of firm h is

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

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

$$-F(\theta^G)C_k(X(k_h, \mathbf{k}^G_{-h}), k_h) - R_k(k_h) + \int_{\underline{\theta}}^{\theta^G} \Psi_{k_h}(k_h, \mathbf{k}^G_{-h}, \theta)dF(\theta)$$

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.18), the host country and the firms all internalize the full economic effects of their actions.

Proposition A.6 *Assume that there are $H \geq 2$ identical foreign firms in the industry and that each firms treats prices, regulation and the environmental impact of the other firms as exogenous to the own investment. Assume also that the operating profit at the efficient outcome is sufficiently large: $\Pi(\mathbf{k}^G) - R(k^G) \geq \max_{k\geq 0}\{F(\theta^G)\hat{\Pi}(P(\mathbf{k}^G), k) - R(k)\}$. In this case, the fully efficient outcome $(\mathbf{k}^G, \theta^G)$ can be implemented as a Nash equilibrium by an international investment agreement stipulating relative compensation according to (A.18).*

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

Implementation of the efficient outcome is independent of any information concerning the extent to which the host country internalizes operating profit. 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 $S(\mathbf{k}^G,\theta)+\sum_{h=1}^{H}[\gamma_h \Pi^h(\mathbf{k}^G)+(1-\gamma_h)T^h(\mathbf{k}^G)]$ at the efficient investment profile $\mathbf{k}^G$. This is equal to $S(\mathbf{k}^G,\theta)+\sum_{h=1}^{H}\Pi^h(\mathbf{k}^G)$ under (A.18) and therefore independent of all γ_h because $T^h(\mathbf{k}^G)=\Pi^h(\mathbf{k}^G)$ for all h.

Proposition A.6 holds for some degree of firm heterogeneity. The important part is that firms are sufficiently similar that $|\mathcal{H}(\mathbf{k}^G)|=H$, so that the ex post incentive to regulate is efficient at $\mathbf{k}^G$. 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 $\mathbf{k}^G$.