ECONOMICS OF INTERNATIONAL INVESTMENT AGREEMENTS[1]

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

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

October 4, 2019

[1]We are grateful to Robert L. Howse, Petros C. Mavroidis, and Damien J. Neven for detailed comments, and to Kyle Bagwell, Espen Moen, Ralph Ossa, and Alan O. Sykes for discussions on the topic of the paper. We are also grateful for comments received during presentations at the IFN conference "Globalization and New Technology: Effects on Firms and Workers;" at the 1st Applecross Workshop in Economics; at the conference "Creative Perspectives on International Trade and Foreign Direct Investment: A Celebration to Honor Jim Markusen" at University of Colorado Boulder; at Columbia Law School Trade Seminar 2018; and at the BI Norwegian Business School. The authors thank Riksbankens Jubileumsfond (Horn), the Marianne and Marcus Wallenberg Foundation (Horn), and Jan Wallanders och Tom Hedelius stiftelse (Tangerås), for financial support.

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

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

Abstract

Nearly 2700 highly potent international investment agreements protect foreign investment against host country policies. This paper analyzes the highly contentious provisions regarding regulatory expropriations that are included in most agreements. These stipulations request compensation for policy interventions other than direct take-overs of assets. In the model countries negotiate the circumstances under which compensation should be paid, and when the host country can regulate without compensation. We show e.g. that: (i) Despite the simplicity of the scheme, it fully mitigates distortions both to foreign investment and to host country regulation under robust circumstances; (ii) Agreements will not cause underregulation from a joint perspective; (iii) Agreements will include investor-state dispute settlement; (iv) North-South agreements solve the inability of South to protect inward investment, while North-North agreements solve Prisoners Dilemma problems with regard to investment protection for reciprocal investment flows, benefitting investors at the expense of the rest of society.

JEL Codes: F21; F23; F53; K33

Keywords: Carve-out compensation; expropriation; foreign investment; international investment agreement; regulatory chill

Contents

1 Introduction

International investment agreements are state-to-state treaties that aim to promote foreign investment by protecting foreign investors against host country policy interventions. The agreements typically require host countries to compensate investors in case of direct expropriation or for measures with similar effects, and they contain a range of other substantive provisions. The agreements also almost invariably include investor-state dispute settlement (ISDS) mechanisms that enable foreign investors to pursue disputes regarding alleged violations of the agreements outside the host country legal systems. There are approximately 2 700 agreements currently in force worldwide.[1] Most of these agreements are bilateral treaties that solely address investment protection, but preferential trade agreements also increasingly often include such stipulations. For instance, investment protection now is a common feature of EU and US preferential trade agreements.

Investment agreements were initially formed without much political opposition, but have more recently become the subject of intensive debate.[2] In developed countries the discussion has mostly concerned investment protection in large agreements such as the *North American Free Trade Agreement* (NAFTA), the *Trans-Pacific Partnership* (TPP), the Canada-EU *Comprehensive Economic and Trade Agreement* (CETA), and the EU-US *Transatlantic Trade and Investment Partnership* (TTIP). The debate has been fuelled by a number of high-profile investment disputes. One example is the $15 billion damage payment sought by TransCanada Corporation for the Obama administration's rejection of the Keystone XL pipe line, a decision that was later overturned by the Trump administration. Other contentious examples are the dispute brought by the energy company Vattenfall against Germany regarding the German decision to accelerate the phase-out of nuclear power in the wake of the Fukushima disaster, and the dispute between Philip Morris and Australia over tobacco plain packaging legislation. A main contention in the debate is that the agreements cause "regulatory chill," that is, they prevent host countries from undertaking desirable regulation. For instance, in the words of US Trade Representative Robert Lighthizer:

> [W]e had situations where real regulation which should be in place, which is bipartisan and in everybody's interest, has not been put in place for fears of ISDS.[3]

The political debate has also concerned the distribution of the benefits and costs of these agreements. Some claim that the agreements only benefit multinational corporations from richer countries, while others argue that the agreements benefit host countries by increasing employment, generating technological transfers, and so forth.

The perceived problems have led to extensive revisions of existing agreements, and to different drafting of newer agreements. A common theme has been to reduce the ambit of central substantive

[1] An extensive list of investment agreements can be found at http://investmentpolicyhub.unctad.org/IIA.

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

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

provisions in order to grant host countries more freedom to regulate without having to compensate investors. For instance, CETA stipulates in Art. 8.9 that:

> ...the Parties reaffirm their right to regulate...to achieve legitimate policy objectives, such as the protection of public health, safety, the environment or public morals, social or consumer protection...,

and similar reservations appear in the *2012 U.S. Model Bilateral Investment Treaty.* Older agreements, many of which are still in force, do not include such explicit reservations. There have also been changes to dispute settlement provisions. For example, Canada has withdrawn completely from ISDS in the revised draft of the NAFTA, and significant limitations have been inserted regarding the possibility for Mexican investors to initiate disputes against the US and vice versa.

Despite the controversies surrounding investment agreements, they have received very little attention in the economic literature (see the review below). This paper contributes to the understanding of these agreements by focusing on a main source of contention: their substantive undertakings concerning *regulatory (or indirect) expropriation.* These provisions stipulate circumstances under which the host country must compensate foreign investors for policy interventions that deprive investors of the return on their investments, but without formal taking of assets by the host country. We examine a number of important questions regarding regulatory expropriation in investment agreements: Do the provisions solve the problems they aim to address? If not, how could they be modified to achieve better outcomes? Do these provisions cause regulatory chill? Why do almost all agreements rely on ISDS? Who benefits and who loses from investment agreements? Why are some agreements more controversial than others? Do agreements between poor and rich countries serve the same purposes as agreements between rich countries?

To address these issues, we lay out a two-country model of investment and regulation. Absent an agreement, firms in a source country decide how much to invest in production facilities in a host country. Production benefits the host country through higher employment, higher incomes, or technology transfers, and so forth. But production can also have some adverse consequence, such as pollution or a health hazard. The magnitude of this regulatory shock becomes known only after the investments have been sunk and can be sufficiently severe to make production undesirable from a host country and even a joint perspective. Having observed the shock, the host country decides whether to permit or to disallow (regulate) production.

Two externalities cause the equilibrium outcome to differ from the investment portfolio and regulation that maximize the expected joint surplus of the two countries. First, investors disregard the external effects of their investments. Second, the host country disregards the losses suffered by the investors in case of regulation. These distortions create scope for an investment agreement.

The countries negotiate a Pareto optimal investment agreement at the outset of the interaction, before investment decisions are made. The agreement specifies how much compensation investors shall receive for regulation in different circumstances.[4] When negotiating the agreement, the source country is only concerned with the expected industry profits of its outward investments and therefore prefers protection to be as broad as possible. The host country also has an interest in increasing investment protection in order to stimulate investment. But the host country will pay for the protection in terms of reduced regulation and/or through compensation payments, and therefore typically prefers less protection than the source country. There is thus a conflict of interest between the parties regarding the design of the agreement that might affect the resulting investments and host country regulatory decisions.

A basic aim of this paper is the focus on contract stipulations that are consistent with those found in actual agreements, to highlight the efficiency properties and consequences of these agreements, and how they could potentially be improved. Virtually all actual agreements share certain core features. First, transfer payments are only requested as compensation for host country policy interventions; the agreements do not stipulate investment-specific subsidies or taxes. Second, all compensation payments go directly from host countries to affected foreign investors. Hence, agreements do not specify compensation payments in the opposite direction from investors to host countries, nor do third parties receive or contribute to such compensation payments. Third, reflecting basic principles in international law, reparations must fully compensate investors for their losses, but they cannot exceed those losses. But there are fundamental differences in the protection the agreements offer investors. Some agreements include a number of explicit carve-outs that allow countries to regulate without paying compensation to achieve certain policy-sensitive objectives; the above quote from CETA is an example. Other agreements have few or no such carve-outs.

To capture these features we consider agreements that specify a threshold for compensation, such that investors receive full compensation for foregone operating profits if regulation takes place for regulatory shocks below the threshold, but no compensation if regulated for shocks above the threshold. This threshold depends on the size of investments because investments affect the value of production relative to regulation. The threshold determines the investment protection provided by the agreement, because investors are guaranteed their operating profits for all shocks below the threshold, regardless of whether they are allowed to produce or are regulated. Despite its apparent restrictions, this *carve-out compensation scheme* will be shown to have some very attractive properties.

Our analysis generates a large number of novel results that yield insights into the design of investment agreements and the validity of arguments in the policy debate. We state the most important ones here.

A first result is that a negotiated agreement based on carve-out compensation can implement the jointly efficient outcome in a non-trivial set of circumstances. This property is quite remarkable

[4] The agreement cannot simply prohibit regulation since regulation is jointly desirable for severe regulatory shocks. This property renders the setting qualitatively different from a standard hold-up problem in which intervention can only reduce joint surplus.

in view of the problems that a negotiated efficient agreement must simultaneously address: It must solve the investment distortions for all firms, it must off-set the host country's incentive to overregulate, and it must distribute the surplus according to the relative bargaining power of the two countries. In other words, carve-out compensation can achieve the same outcome in terms of efficiency and distribution of the gains from the agreement as much more complicated compensation schemes. Hence, two simple features of actual agreements—the "all-or-nothing" principle by which investors are either fully compensated or not compensated at all, and the right to regulate without compensation in certain circumstances—are not as restrictive as they might seem.

A second result of relevance for actual agreements is that implementation of the jointly efficient outcome requires compensation payments to occur in equilibrium for certain shocks. This suggests that such payments neither reflect excessive litigation by private investors, nor jointly undesirable regulation by the host country, contrary to common claims in the debate. The payments instead serve as efficient implicit investment subsidies. They also function as implicit side-payments in the negotiation of the investment agreement that allow the parties to distribute the surplus of the agreement without distortionary effects. Hence, investment agreements, if properly designed, can fully replace investment subsidies.

A third result is that countries will negotiate carve-out compensation as long as compensation cannot exceed foregone operating profits. This holds even if carve-out compensation cannot achieve joint efficiency. To improve on the efficiency of the outcome it is necessary to introduce contractual features not found in actual agreements. We point to ways in which this can be done.

A fourth result is that the countries may negotiate a reciprocal agreement, by which the same conditions apply to both signatory countries, even they are asymmetric. Reciprocity is immaterial if investments flow in one direction only, as in our baseline case. But negotiations can yield a jointly efficient agreement that applies identically to both parties also when investments flow in both directions between the countries. This finding sheds light on the reciprocal nature of actual investment agreements.

We then extend the analysis of our baseline model by considering three other standard provisions in investment agreements. In a first extension we introduce a *National Treatment* (NT) provision. It is sometimes argued that investment agreements should go no further than to provide for non-discriminatory treatment of inward investment. We find no support for this recommendation in our setting. On the contrary, the negotiated agreement will never consist of an NT provision only, as long as carve-out compensation can be included in the agreement. If the Pareto optimal agreement also includes NT, the role of the provision is to enable a host country to use the enforcement mechanism in the agreement to indirectly correct distorted incentives for domestic investors. This finding is consistent with the fact that agreements very rarely protect against non-discriminatory treatment of inward investment only.

The second extension is to consider compensation rules for *direct expropriation,* that is, for cases where a host country seizes investors' assets. Actual agreements sometimes specify the same reparations for direct and regulatory expropriation, and sometimes provide complete investment protection for direct expropriation only. We demonstrate that negotiated agreements do not include any carve-outs for direct expropriation if assets are of sufficiently small value to the host country when expropriated. In the opposite case, the negotiated investment protection is the same for both types of expropriation.

The third extension concerns the *dispute settlement system*. A commonly suggested reason for the inclusion of ISDS in investment agreements is that state-state dispute settlement (SSDS) causes diplomatic or political frictions. On the one hand, these frictions are costly and lead to weak incentives for source country governments to implement the agreement. But on the other hand, ISDS can yield excessive investor incentives to request arbitration. To shed some light on the choice between ISDS and SSDS, we assume that SSDS is associated with dispute settlement costs that do not arise under ISDS. We show that it might be in the host country's unilateral interest to rely exclusively on SSDS, but that this would reduce joint surplus. The negotiated agreement will therefore include ISDS. This result is consistent with the fact that the vast majority of investment agreements feature ISDS.

Our final set of findings concerns policy aspects of investment agreements. A core issue is whether agreements cause regulatory chill. We distinguish between domestic regulatory chill, which occurs when a lower propensity to regulate reduces the host country surplus, and joint chill, which occurs when instead the joint surplus is reduced. Any agreement will cause domestic regulatory chill since a main purpose is to incentivize investment precisely by reducing regulation. We also establish that equilibrium investment agreements will not yield underregulation from a joint perspective under standard circumstances, but provide an example of lobbying that gives rise to joint chill.

A second key policy question concerns the distributional effects of the agreements. We show that these implications depend on two fundamental features of the contracting situation. The first is *whether investment flows are one-way or two-way*. The other is the ability of the host country (or countries) to make *credible unilateral commitments* to protect inward investment absent an agreement. Our baseline agreement covers one-way investment from a source country to a host country that lacks the ability to credibly commit to investment protection. This illustrates the setting for a traditional bilateral investment treaty between a developed country (North) and a developing country (South). The negotiated agreement will benefit investors in North and increase surplus in South by the assumption that participation in the agreement is voluntary. The context of a North-North agreement between two developed economies is very different. Agreements such as CETA or TTIP intend to stimulate investment in both directions, and the economies can credibly commit to protect inward investment through their constitutions, laws and regulations even if there is no agreement. In our North-North setting, countries would unilaterally set investment protection to maximize national surplus absent an agreement. But countries care also about the expected profits of outward investment when they negotiate the agreement. Because investors benefit from increased investment protection, these countries will negotiate investment protection beyond the levels that maximize national surplus. We thus find that the North-North agreement benefits investors at the expense of the rest of society. This finding is consistent with the strong public opposition to agreements such as TTIP.

Our final observation concerns the role of investment agreements. The purpose of a North-South agreement is to help South solve a *hold-up problem* stemming from overregulation. This is achieved by way of the third-party enforcement mechanisms that underlie the international investment agreement regime. The partners to a North-North agreement do not require this help. The problem they face is instead that their unilateral decisions on investment protection disregard the benefits to foreign investors. The purpose of a North-North agreement is thus to solve a *Prisoner's Dilemma problem* by committing both parties to give more protection to foreign investment than is unilaterally optimal.

Contribution to the literature Investment agreements became the subject of formal economic analysis in the late 1990s, mainly inspired by the failed attempt by the OECD to launch its *Multilateral Agreement on Investment* (e.g. Markusen, 1998, 2001; Turrini and Urban, 2008).[5] Attention has also been devoted the relationship between preferential trade agreements and investment agreements (e.g. Bergstrand and Egger, 2013). But until recently the dominant theme in the small economic literature has been the extent to which investment agreements stimulate investment in practice (see Falvey and Foster-McGregor, 2017, for a recent contribution).

A nascent theoretical literature addresses issues closer to the current policy debate. One approach is to examine implications of exogenously specified agreements. Konrad (2017) and Schjelderup and Stähler (2019) show how investment agreements might induce strategic overinvestment by foreign firms. Janeba (2019) formally defines the amorphous notion of regulatory chill and examines its occurrence in a specific setting. Kohler and Stähler (2019) compare an investment agreement that provides compensation when regulatory policies are changed in unfavorable direction for investors, with an agreement that instead comprises a National Treatment provision. A number of interesting observations emerge from these papers. Most striking is the finding that investment agreements can reduce the surplus of host countries, and sometimes even joint surplus. But this finding also points to the difficulty of using an approach that imposes an exogenously specified agreement on the parties. The existence of thousands of investment agreements raises the question whether it is plausible that countries at such a large scale have entered into agreements that are not in their own long-run interest.

Closer to this paper is a second line of investigation that analyzes the design of efficient agree-

[5] A earlier and mostly informal literature studies direct expropriation of foreign investment absent investment agreements; a well-known example is Vernon's (1971) "obsolescing bargaining" theory. Some formal studies focuses on how reputation mechanisms can remedy investor-host country hold-up problems; see for instance Dixit (1988), and Thomas and Worral (1994). Dixit (2011) reviews the literature and discusses a range of issues related to insecurity of property rights and foreign investment.

ments. Aisbett, Karp and McAusland (2010a) show how a carve-out scheme under which investors receive compensation in excess of foregone operating profits, can achieve an efficient outcome in a model with distorted incentives to regulate and where arbitration courts are imperfectly informed about the magnitude of regulatory shocks.[6] Aisbett, Karp and McAusland (2010b) highlight the interaction between National Treatment provisions and compensation requirements, under the assumption that the host country can charge investment-specific payments for investment protection. Stähler (2018) draws on mechanism design to characterize an efficient compensation mechanism where the payment balance between the host country and investors is broken and where compensation is based on host country utility of regulation rather than foregone operating profits. The compensation schemes in all three papers have properties—excess compensation, investment-specific taxes, third-party transfers—that are not found in actual agreements. It is therefore difficult to draw conclusions about the properties of actual agreements on the basis of these papers. Ossa, Staiger and Sykes (2019) take a different approach by analyzing the efficient choice of dispute settlement stipulations among a set of contract provisions found in actual trade and investment agreements. They address issues that are largely complementary to those studied here, however, although there is some overlap concerning the analysis of ISDS versus SSDS; see Section 5.3 for a detailed discussion.

Our analysis focuses on endogenously and purposefully designed agreements, similar to the four papers just mentioned. But in contrast to these papers, and the rest of the literature, we develop a descriptive theory of investment agreements in which countries *negotiate contractual instruments similar to those found in actual agreements.* The paper is also unique in distinguishing between agreements with one-way and two-way investment flows, in examining how countries' abilities to make unilateral commitments regarding investment protection affect the negotiated outcomes, and by investigating the interaction between rules for regulatory expropriation with other core undertakings in actual agreements. In sum, we identify factors that determine the efficiency of agreements, the distribution of the surplus, the scope for entering into the agreements, and the nature of the problems investment agreements address.

The rest of the paper is organized as follows. Section 2 lays out the economic setting. Section 3 introduces our formalization of an investment agreement. Section 4 characterizes the negotiated agreement, and makes a number of observations concerning its features. Section 5 analyzes the extensions to NT, direct expropriation and SSDS. Section 6 discusses policy issues. Section 7 concludes and contains a list of suggestions for future research. All formal proofs appear in the Appendix.

[6] Miceli and Segerson (1994) introduce carve-out compensation in their study of the limit of a government's right to regulate private property. They demonstrate the efficiency of this scheme in a model where incentives to invest and regulate are undistorted. Under those assumptions it is also optimal not to compensate regulatory takings (Blume, Rubinfeld and Shapiro, 1984). Hermalin (1995) derives two efficient compensation mechanisms for direct expropriation. In the first, the investor pays a production tax equal to the value of seizing the asset. In the second, the government pays the value of seizing the asset as compensation. This second compensation rule is efficient under direct, but not regulatory expropriation.

2 The problems for an investment agreement to solve

This section lays out the model, identifies the two fundamental distortions that investment agreements seek to address, and characterizes the equilibrium absent investment protection.

2.1 The economic environment

There are two countries, "Home" and "Foreign". Their only economic interaction is through investment by Foreign firms in Home.[7] There is a single industry with $H \geq 1$ Foreign risk-neutral firms.[8] At the outset of the interaction, each firm $h \in \{1, ..., H\} \equiv \mathcal{H}$ decides how much capital $k_h \geq 0$ to invest in Home. Let $\mathbf{k} = (k_1, ..., k_H)$ be the portfolio of foreign investment. These investments are irreversible. Firm h's investment cost is given by the continuous and strictly increasing function $R^h(k_h)$, where $R^h(0) = 0$. The firm's operating profit $\Pi^h(\mathbf{k})$ is continuous and strictly positive for $k_h > 0$, but zero for $k_h = 0$. Let $R(\mathbf{k}) \equiv \sum_{h=1}^{H} R^h(k_h)$ be the industry investment cost, and let $\Pi(\mathbf{k}) \equiv \sum_{h=1}^{H} \Pi^h(\mathbf{k})$ be the industry operating profit.

Foreign investment benefits Home, the host country, by creating consumer surplus, employment, technological spill-overs, learning-by-doing by the work-force, and so forth. But the investment can also have some adverse consequence, such as pollution or health hazard. To represent this, we assume that an industry-specific shock θ that affects the utility of Home is realized after foreign investment is sunk. A large value of θ means a more negative shock, which could represent the arrival of severely adverse information regarding environmental or health consequences of the production process or the goods produced, or other factors that significantly reduce the desirability of the investment. We will not adopt any particular interpretation, but simply denote θ as capturing a "regulatory shock." All externalities from the investments arise during the production stage, and they appear only in case of production. Ex ante, the shock is continuously distributed on $[\underline{\theta}, \bar{\theta}]$ with cumulative distribution function $F(\theta)$ and density $f(\theta)$.

In the final stage of the interaction, having observed $\mathbf{k}$ and θ, the host country decides whether to permit production by all H firms, or to regulate by disallowing production in the whole industry. Let $V(\mathbf{k}, \theta)$ denote the host country utility if there is production. We do not make any specific interpretation of the objective function $V(\mathbf{k}, \theta)$ other than to assume that it is consistent over time. The marginal utility of investment by firm h can be positive or negative, $V_h(\mathbf{k}, \theta) \gtreqless 0$ (subscripts on functional operators denote partial derivatives throughout, and subscript h denotes the partial derivative with respect to k_h). As stated above, a larger shock θ reduces the utility of allowing production, $V_\theta(\mathbf{k}, \theta) < 0$. To avoid less interesting corner solutions, we assume for all $\mathbf{k} \neq \mathbf{0}$ that host country utility is non-negative at the most favorable realization of the shock, $V(\mathbf{k}, \underline{\theta}) \geq 0$, and

[7] Two-way investment is qualitatively different. We consider this case in Section 4.4.

[8] A defining characteristic of international investment agreements is their economy-wide scope. We extend the model to cover multiple industries in Section 4.5. This extension has some interesting implications regarding efficiency, but does not fundamentally change the analysis. We allow for domestic firms in the analysis of National Treatment in Section 5.1, but otherwise assume away domestic firms to simplify the exposition.

negative at the most unfavorable realization, $V(\mathbf{k},\bar{\theta}) < 0$.

Our model allows different firms' investments to have different effects on host country utility $V(\mathbf{k},\theta)$. In general, it could be in the host country's interest to regulate only a subset of firms. For the most part of the analysis, we take our definition of an industry to mean a set of firms that contribute in a sufficiently similar manner to Home's utility that Home either allows production by all firms or regulates the whole industry. Hence, $V(\mathbf{k},\theta)$ is realized and every firm $h \in \mathcal{H}$ receives its operating profit $\Pi^h(\mathbf{k})$ if the host country allows production. In case of regulation, the host country utility is zero, and all H firms receive zero operating profits. We consider asymmetric regulation in Section 4.5 where we examine multiple industries, and in Section 5.1 where we analyze National Treatment provisions.

We derive the equilibrium outcome in standard fashion throughout, by solving for the interaction backwards, starting with the regulatory decision.

2.2 Regulation

When deciding whether to permit production or to regulate, the host country considers the implications for its own utility $V(\mathbf{k},\theta)$ of foreign investment, but disregards the loss of operating profits $\Pi(\mathbf{k})$ for regulated firms (which is the sole consequence of regulation for Foreign interests). For investment profile $\mathbf{k}$, the host country is indifferent between allowing all firms to produce, or regulating all firms if $V(\mathbf{k},\theta) = 0$. The critical level of the regulatory shock $\Theta(\mathbf{k}) \in [\underline{\theta},\bar{\theta})$ is defined by

$$V(\mathbf{k},\Theta(\mathbf{k})) \equiv 0.$$

It is ex post optimal for the host country to allow production by all firms for $\theta \leq \Theta(\mathbf{k})$, and to regulate all firms for $\theta > \Theta(\mathbf{k})$, because $V(\mathbf{k},\theta)$ is strictly decreasing in θ. We assume that the host country allows production if indifferent.

2.3 Investment

Firms make their investment decisions simultaneously and independently to maximize their expected profits. Let $\mathbf{k}^0 = (k_1^0, ..., k_H^0)$ be an equilibrium vector of foreign investment. We do not make any assumptions regarding the nature of strategic interaction at the investment stage, but firms rationally foresee the consequences of their respective investment on regulation and incorporate such effects into their investment decisions. Hence, each firm h will in a (subgame-perfect) equilibrium invest

$$k_h^0 \in \arg\max_{k_h \geq 0} \{F(\Theta(k_h, \mathbf{k}_{-h}^0))\Pi^h(k_h, \mathbf{k}_{-h}^0) - R^h(k_h)\},$$

where $\mathbf{k}_{-h}^0$ constitutes the equilibrium investment portfolio of all firms other than h.

Let v^0 be the expected host country utility, and let π^0 be the expected industry profit in equilibrium,

$$v^0 \equiv \int_{\underline{\theta}}^{\theta^0} V(\mathbf{k}^0, \theta) dF(\theta),\ \pi^0 \equiv F(\theta^0)\Pi(\mathbf{k}^0) - R(\mathbf{k}^0), \tag{1}$$

where $\theta^0 = \Theta(\mathbf{k}^0)$ is the equilibrium threshold for regulation.

2.4 The general inefficiency of investment and regulatory decisions

We will use the joint surplus $V(\mathbf{k}, \theta) + \Pi(\mathbf{k})$ of the two countries as a benchmark for evaluating the efficiency of regulation, for reasons to be explained in Section 4. The ex post jointly optimal threshold for regulation $\Theta^J(\mathbf{k}) > \underline{\theta}$ is given by

$$V(\mathbf{k},\Theta^J(\mathbf{k})) + \Pi(\mathbf{k}) \equiv 0$$

if $V(\mathbf{k},\bar{\theta}) + \Pi(\mathbf{k}) \leq 0$ and by $\Theta^J(\mathbf{k}) = \bar{\theta}$ otherwise. It follows from $\Pi(\mathbf{k}) > 0$ and $V_\theta(\mathbf{k},\theta) < 0$ that $\Theta^J(\mathbf{k}) > \Theta(\mathbf{k})$. Consequently:

Observation 1 *Assume that there is no investment protection. From a joint surplus perspective, for any arbitrary investment profile* $\mathbf{k}$*, the host country:*
(i) correctly allows production for $\theta \leq \Theta(\mathbf{k})$*;*
(ii) overregulates for $\theta \in (\Theta(\mathbf{k}), \Theta^J(\mathbf{k}))$*; and*
(iii) correctly regulates for $\theta \geq \Theta^J(\mathbf{k})$*.*

Under ex post optimal regulation, the expected joint surplus of the host country and foreign firms equals

$$\int_{\underline{\theta}}^{\Theta^J(\mathbf{k})} [V(\mathbf{k},\theta) + \Pi(\mathbf{k})] dF(\theta) - R(\mathbf{k}). \tag{2}$$

The jointly optimal investment k_h^J by firm h is thus given by

$$\int_{\underline{\theta}}^{\theta^J} V_h(\mathbf{k}^J, \theta) dF(\theta) + F(\theta^J)\Pi_h(\mathbf{k}^J) - R_h^h(k_h^J) = 0$$

at an interior optimum $k_h^J > 0$, where $\theta^J \equiv \Theta^J(\mathbf{k}^J)$ is the ex post optimal threshold for regulation at the jointly optimal investment portfolio $\mathbf{k}^J = (k_1^J, ..., k_H^J)$. The first term in the above expression is the externality of h's investment on the host country, the second is the marginal effect on industry profit, including the investment externality on all firms other than h, and the last term is the marginal investment cost to the firm.

There are several reasons why the equilibrium outcome $(\mathbf{k}^0, \theta^0)$ generally differs from the jointly optimal outcome $(\mathbf{k}^J, \theta^J)$. First, investors expose the host country to externalities from their investments; second, the host country exposes investors to externalities when regulating the industry; and third, investors expose each other to externalities from the investment decisions. We have not imposed enough structure to unambiguously determine the aggregate impact of these externalities.

One can show that the equilibrium $(\mathbf{k}^0, \theta^0)$ can feature simultaneous overregulation and underinvestment, which would constitute the type of problems that investment agreements typically are meant to address.[9] But the results to follow do not hinge on the equilibria having this particular feature.

3 A framework for analyzing regulatory expropriation provisions

State-to-state investment agreements are long-term commitments to protect foreign investment against host country policy interventions. This section lays out our formalization of the obligations to compensate investors for *regulatory (indirect) expropriations* that are almost invariably included in these agreements. The obligations mandate compensation for host country measures that deprive investors of the return on their investments without involving formal seizure of assets.[10] We discuss other central provisions in Section 5.

An investment agreement states how much compensation foreign investors should receive in case of regulatory intervention. Because of their scope, actual agreements lay down general rules for compensation instead of specifying damage payments to individual firms (this need not be the case in commercial contracts). Yet, asymmetries across firms imply that some investors can receive more compensation than others even if they are subject to the same compensation rule. A relevant example is a rule which specifies that compensation should equal foregone operating profit. To allow for such asymmetries, we let compensation T^h to each firm be indexed by h, and denote a compensation scheme as a vector $\mathbf{T} \equiv (T^1, ..., T^H)$. The parties might in general benefit from conditioning compensation payments on any pay-off relevant information. Since the investment portfolio $\mathbf{k}$ and the magnitude of the regulatory shock θ completely describe the host country utility and investor profits, the compensation function $T^h = T^h(\mathbf{k}, \theta)$ for each firm h encompasses any relevant compensation scheme.

3.1 The sequence of events

We assume that events unfold as follows:

1. Home and Foreign negotiate a binding agreement on a compensation rule $\mathbf{T}$ for regulation of investments undertaken by Foreign firms in Home.

2. Each firm $h \in \mathcal{H}$ makes an irreversible investment k_h.

[9] In Section 5.2, we consider the case where the host country can intervene either by regulation or direct expropriation by which it seizes the firms' assets. Absent investment protection, it will be optimal for the host country to do one or the other after the investment has been sunk, depending on the magnitude of θ. Either way, firms receive zero operating profit no matter how much they invest, in which case $\mathbf{k}^0 = \mathbf{0}$.

[10] According to case law, measures must deprive investors of almost all their profits in order to possibly constitute indirect expropriation. See Dolzer and Schreuer (2012) for a comprehensive overview of International Investment Law.

3. The regulatory shock θ is observed; and
4. Home chooses whether to:
   (a) Permit production, in which case Home utility is $V(\mathbf{k}, \theta)$ and each firm h receives $\Pi^h(\mathbf{k})$, or:
   (b) Regulate and pay compensation, in which case Home utility is $-\sum_{h=1}^{H} T^h(\mathbf{k}, \theta)$, and each firm h receives $T^h(\mathbf{k}, \theta)$.

We refer to stages 2-4 as the *market game* induced by the agreement with compensation rule $\mathbf{T}$.

An important feature of international investment agreements is their highly potent enforcement mechanisms. Agreements commonly build on international conventions that require courts in the signatory nation states to recognize and enforce arbitral awards from any other signatory nation state.[11] Hence, investment agreements rely on a form of mandatory third-party enforcement, and have in this regard much stronger enforcement mechanisms than trade agreements. In accordance with much of the literature on trade agreements, we assume that the agreement is costlessly enforceable, and that compensation is simply paid without any need for a formal arbitration stage of the game.[12]

Decisions to regulate are irreversible in our model. There are many real-life examples when temporary regulation makes the re-opening of operations effectively infeasible, even if not legally so. For instance, the time lag between regulation and an arbitration panel's decision to overturn a host country intervention can render a production plant obsolete. Or the host country can simply have demolished the plant in the meantime. Yet, adding the possibility of re-opening plants would be without consequence within our model because it would always be better for the host country to pay compensation compared to re-allowing production, due to perfect contract enforcement; we will return to this below.

3.2 Compensation schemes

We want to restrict the structure of the agreement under consideration to reflect the compensation mechanisms in actual agreements, although we will also examine the extent to which these features constrain the outcome. A first restriction is that the agreement *does not allow payments from firms to the host country*. Hence, $T^h(\mathbf{k}, \theta) \geq 0$. A second restriction is that there is *no third-party involvement* in compensation payments. These features are captured in the following definition:

[11] The two main conventions are the UN-based "New York Convention" with 157 members, and the World Bank Group-based "ICSID Convention" with 159 contracting states.

[12] Fundamental results of our paper do not depend crucially on such perfect enforcement. We discuss implications of imperfect enforcement in Section 4.2 and elsewhere.

Definition 1 *A **general compensation scheme** is a vector of transfers* $\mathbf{T} = (T^1, ..., T^H)$*, where* $T^h(\mathbf{k}, \theta) \geq 0$ *for all* h*, to be paid if and only if the host country regulates, and without third-party involvement.*

The general compensation scheme does not limit the amount of compensation, nor does it specify the type of situation in which compensation is to be paid, and it therefore allows for a range of features not found in actual agreements. Similar to Miceli and Segerson (1994), and Aisbett, Karp and McAusland (2010a), we focus on a much more specific type of compensation:

Definition 2 *A **carve-out compensation scheme** is a vector of transfers* $\mathbf{T}^C = (T^{1C}, ..., T^{HC})$*, where*

$$T^{hC}(\mathbf{k}, \theta) \equiv \begin{cases} \Pi^h(\mathbf{k}) & \text{if } \theta \leq \Theta^C(\mathbf{k}) \\ 0 & \text{if } \theta > \Theta^C(\mathbf{k}) \end{cases} \quad \text{for all } h, \tag{3}$$

to be paid if and only if the host country regulates, and without third-party involvement.

The carve-out compensation scheme $\mathbf{T}^C$ is a special case of the general compensation scheme, only it requires the host country to pay compensation for the full foregone operating profit to each firm h if regulation occurs for shocks below a threshold $\Theta^C(\mathbf{k})$, but allows regulation without compensation payments for more severe shocks. We refer to the threshold $\Theta^C(\mathbf{k})$ as the *investment protection* provided by $\mathbf{T}^C$, since investors receive $\Pi^h(\mathbf{k})$ for all shocks $\theta \leq \Theta^C(\mathbf{k})$ regardless of whether they are allowed to produce or are regulated.

The carve-out compensation scheme reflects actual agreements and their interpretations in several important regards. First, it requests investors to be compensated with their respective *full foregone operating profits* whenever compensation is due. This is standard practice in actual investment disputes, and it reflects a fundamental principle in international law concerning state responsibility, which holds that "...reparation must, as far as possible, wipe out all the consequences of the illegal act and re-establish the situation which would, in all probability, have existed if that act had not been committed...."[13]

Second, while traditional agreements contain very few, if any, explicit exceptions for regulatory policies, there is a strong tendency to include *carve-out provisions* in new or revised agreements. For instance, the *2012 U.S. Model Bilateral Investment Treaty* establishes that "[e]xcept in rare circumstances, non-discriminatory regulatory actions by a Party that are designed and applied to protect legitimate public policy objectives, such as public health, safety and the environment, do not constitute indirect expropriations." The concept of a legitimate policy intervention has an intuitive interpretation under compensation $\mathbf{T}^C$: Such interventions occur for the subset of shocks $\theta \in (\Theta^C(\mathbf{k}), \bar{\theta}]$ of sufficient magnitude that the host country does not have to pay compensation for

[13] This often cited quote is from the ruling by the Permanent Court of International Justices (the predecessor to the International Court of Justice) in the *The Factory at Chorzów* case from 1928. Another central notion is that "[t]he compensation shall cover any financially assessable damage including loss of profits insofar as it is established." (Article 36, International Law Commission, 2001, with a footnote omitted).

regulation. $\mathbf{T}^C$ thus has a carve-out for the interval $(\Theta^C(\mathbf{k}), \bar{\theta}]$. A fundamental policy question is how well carve-out compensation schemes can solve inefficiency problems associated with foreign investment and regulatory intervention.

Third, the carve-out compensation scheme *does not allow punitive payments* in the sense of requesting host countries to pay larger compensation than the harm suffered by investors. This important principle in international law is increasingly often explicitly stated in investment agreements.[14] We will discuss some implications of this restriction below.

Observe that the same carve-out compensation rule $\Theta^C(\mathbf{k})$ applies to all H firms in the industry. Setting the same terms for all foreign investment in a state-to-state agreement is likely to reduce transaction costs relative to a situation in which each firm unilaterally negotiated an investor-state agreement with the host country. Differences in transaction costs can be one reason why investment agreements are signed between countries instead of at a more disaggregated level.[15] In this paper, we simply assume that investment agreements are state-to-state treaties, and set the transaction costs of such agreements to zero.

3.3 The equilibrium in the market game

A subgame-perfect equilibrium of the market game induced by an agreement with general compensation scheme $\mathbf{T}$ consists of two components. First, for any investment profile $\mathbf{k} \neq \mathbf{0}$, the equilibrium defines two subsets of shock realizations, the set $M(\mathbf{k})$ for which it is ex post optimal for the host country to allow production and the complementary set $M^r(\mathbf{k})$ for which the host country optimally regulates. Compensation payments reduce host country intervention relative to the case of no agreement, since the host country allows production for all θ such that

$$V(\mathbf{k}, \theta) + \sum_{h=1}^{H} T^h(\mathbf{k}, \theta) \geq 0$$

and regulates otherwise.[16]

The second component of the equilibrium is an investment profile $\hat{\mathbf{k}} = (\hat{k}_1, ..., \hat{k}_H)$ such that $\hat{k}_h$ maximizes the expected investment profit

$$\hat{k}_h \in \arg\max_{k_h \geq 0} \{ \textstyle\int_{M(k_h, \hat{\mathbf{k}}_{-h})} dF(\theta) \Pi^h(k_h, \hat{\mathbf{k}}_{-h}) + \int_{M^r(k_h, \hat{\mathbf{k}}_{-h})} T^h(k_h, \hat{\mathbf{k}}_{-h}, \theta) dF(\theta) - R^h(k_h) \}$$

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

[15] In Section 4.5, we demonstrate additional benefits of writing an economy-wide agreement.

[16] We here see why the host country would pay compensation subsequent to regulation rather than overturn its initial decision if forced to choose. The latter possibility, which Ossa, Staiger and Sykes (2019) refer to as "cease and desist," yields host country utility $V(\mathbf{k}, \theta)$, while total compensation payments are $\sum_{h=1}^{H} T^h(\mathbf{k}, \theta)$. The host country prefers compensation to cease and desist by construction of $M^r(\mathbf{k})$. By implication, cease and desist can be optimal under an investment agreement only if contract enforcement is imperfect. See Ossa, Staiger and Sykes (2019) for a detailed analysis.

for each firm h given the equilibrium investment profile $\hat{\mathbf{k}}_{-h}$ of all firms except h. In this expression, the first term captures realizations of θ for which there will be no regulation, and the second term those where regulation will occur. The equilibrium investment portfolio $\hat{\mathbf{k}}$ and realizations $M(\hat{\mathbf{k}})$ and $M^r(\hat{\mathbf{k}})$ are all functions of the compensation scheme $\mathbf{T}$, but we subsume $\mathbf{T}$ for notational simplicity.

4 Properties of negotiated regulatory expropriation provisions

The compensation scheme $\mathbf{T}$ is determined through negotiations between the host country and the source country. The negotiating parties correctly anticipate the effects of the agreement on future investment and regulation. We consider first the setting laid out above, where Home and Foreign negotiate an agreement covering investment from Foreign to Home only. This is the typical situation when investment agreements are negotiated between a developed and a developing country, since although formally reciprocal, the agreements effectively only apply to investment from the developed to the developing country. Section 4.4 examines implications of two-way investment flows. We extend the analysis to multiple industries in Section 4.5.

4.1 Bargaining

Consider the negotiations over an agreement in the setting laid out in Sections 2 and 3, with investment potentially flowing from Foreign to Home. The expected utility of Home equals

$$\tilde{V}(\mathbf{T}) \equiv \int_{M(\hat{\mathbf{k}})} V(\hat{\mathbf{k}},\theta)dF(\theta) - \int_{M^r(\hat{\mathbf{k}})} \sum_{h=1}^{H} T^h(\hat{\mathbf{k}},\theta)dF(\theta) \tag{4}$$

under an agreement based on general compensation $\mathbf{T}$. We assume that Foreign's expected utility is the expected investment profit $\tilde{\Pi}(\mathbf{T}) \equiv \sum_{h=1}^{H} \tilde{\Pi}^h(\mathbf{T})$, where the equilibrium expected investment profit of firm h equals

$$\tilde{\Pi}^h(\mathbf{T}) \equiv \int_{M(\hat{\mathbf{k}})} dF(\theta)\Pi^h(\hat{\mathbf{k}}) + \int_{M^r(\hat{\mathbf{k}})} T^h(\hat{\mathbf{k}},\theta)dF(\theta) - R^h(\hat{k}_h). \tag{5}$$

Looking at $\tilde{V}(\mathbf{T})$ and $\tilde{\Pi}^h(\mathbf{T})$, it is easy to understand why investment agreements would be controversial in the host country, but more popular among investors. For given investment $\hat{\mathbf{k}} = \mathbf{k}^0$, the host country can only lose from an agreement because it then allows incremental production precisely in those circumstances under which it would have been be unilaterally optimal to intervene absent an agreement (i.e. for shock realizations $V(\mathbf{k}^0,\theta) < 0$). Moreover, the agreement might also request the host country to pay compensation for regulation. The only reason for the host country to enter into an agreement is thus to affect investment. These benefits only arise when there is no regulation, and can be difficult to quantify. Investors directly benefit from less regulation and from the prospect of compensation payments, even if the agreement has no incremental effect on investment.

With unconstrained side payments, the parties would agree on a compensation scheme that maximizes their joint expected surplus $\tilde{V}(\mathbf{T}) + \tilde{\Pi}(\mathbf{T})$, and then negotiate the side payments to determine the distribution of the surplus. Because the marginal utility of compensation payments is constant and the same for the two contracting parties, the joint surplus depends only indirectly on $\mathbf{T}$ through investment and regulation. Hence, with unconstrained side payments, the agreement would implement the *jointly efficient market outcome* ($\mathbf{k}^J$,θ^J) characterized in Section 2.4, which we use as the benchmark for evaluating the efficiency of agreements. But in keeping with how investment treaties normally are negotiated, we assume that the parties do *not have access to side payments.* This means that joint surplus maximization does not trivially follow from bargaining.[17]

To get a specific prediction for the distribution of the surplus from the agreement, the outcome of the negotiations is taken to be given by the Nash Bargaining Solution, with $v^0 \geq 0$ and $\pi^0 \geq 0$ characterized in (1) as the status quo points for the host and source country. The outcome thus maximizes the Nash Product

$$\mathcal{N}(\mathbf{T}) \equiv [\tilde{V}(\mathbf{T}) - v^0]^\alpha [\tilde{\Pi}(\mathbf{T}) - \pi^0]^{1-\alpha}, \tag{6}$$

where the parameter $\alpha \in (0,1)$ in standard fashion captures the host country "bargaining power" relative to that of the source country.

4.2 When carve-out compensation implements the efficient outcome

The following Proposition shows that the simple compensation scheme $\mathbf{T}^C$, under which firms receive full compensation for regulation below a threshold $\Theta^C(\mathbf{k})$ and no compensation above this threshold, has a remarkable property:

Proposition 1 *A carve-out compensation scheme $\mathbf{T}^C$ implements the jointly efficient market outcome ($\mathbf{k}^J$,θ^J) and maximizes the Nash Product $\mathcal{N}(\mathbf{T})$ in a robust set of circumstances.*

Appendix A.1 verifies a more precisely formulated version of this statement. It thus establishes that for a non-trivial set of circumstances the carve-out compensation scheme *(i)* corrects all firms' investment externalities; *(ii)* corrects the host country regulatory externality, and *(iii)* distributes the surplus of the agreement across the two contracting parties as proposed by the Nash Bargaining Solution. The optimality of $\mathbf{T}^C$ actually holds more broadly than for general compensation schemes $\mathbf{T}$. For instance, one can allow negative compensation.

In what follows, we will first explain the economic mechanisms behind this result, and then draw several implications regarding the economic desirability of the structure of actual agreements.

[17] Contemporary investment protection is often negotiated jointly with other undertakings, such as trade liberalization. This broader scope of the negotiations might open up possibilities for more efficient solutions than in an agreement comprising only investment protection.

4.2.1 How full efficiency and surplus distribution is achieved

We will first show that the incentives to regulate and invest are efficient for all investment protection above a threshold. This can leave enough flexibility to distribute surplus across the two negotiating parties according to the Nash Bargaining Solution.

Regulation incentives Let $\theta^{JC} \equiv \Theta^C(\mathbf{k}^J)$ denote the level of investment protection provided by the carve-out compensation agreement $\mathbf{T}^C$ under efficient investment $\mathbf{k}^J$. If $\theta^{JC} \geq \theta^J$, the host country allows production for $\theta \leq \theta^J$, it regulates for $\theta \in (\theta^J, \theta^{JC}]$ despite having to compensate investors, and it regulates without compensation for $\theta > \theta^{JC}$, all of which is jointly efficient. Hence, the full compensation requirement has the virtue of inducing the host country to *fully internalize* the consequences for investors of its regulatory decisions for any $\theta^{JC} \geq \theta^J$. If instead $\theta^{JC} < \theta^J$, the host country correctly allows production for $\theta \leq \theta^{JC}$, *overregulates* for $\theta \in (\theta^{JC}, \theta^J)$, and correctly regulates for $\theta > \theta^J$. Hence, $\theta^{JC} \geq \theta^J$ is necessary and sufficient for implementing jointly efficient regulation under efficient investment.

Investment incentives A necessary condition for the investment profile $\mathbf{k}^J$ to be an equilibrium under $\mathbf{T}^C$ is that

$$F(\theta^{JC})\Pi^h(\mathbf{k}^J) - R^h(k_h^J) \geq \max_{k_h \geq 0}\{F(\Theta(k_h, \mathbf{k}_{-h}^J))\Pi^h(k_h, \mathbf{k}_{-h}^J) - R^h(k_h)\} \text{ for all } h \in \mathcal{H}. \quad (7)$$

The left-hand side of the inequality is the expected profit of firm h under $\mathbf{T}^C$ if all firms invest $\mathbf{k}^J$. The right-hand side is the expected profit of firm h if it deviates to its optimal investment when there is no investment protection and the other firms invest $\mathbf{k}_{-h}^J$. Since an agreement increases the expected profit for any investment portfolio compared to the case of no investment protection, condition (7) is clearly necessary for $\mathbf{k}^J$ to be an equilibrium.

Inequality (7) is also a sufficient condition for $\mathbf{k}^J$ to be an equilibrium under $\mathbf{T}^C$. To see this, consider for instance the case where investment protection in $\mathbf{T}^C$ is designed such that $\Theta^C(\mathbf{k}) = \Theta(\mathbf{k})$ for all $\mathbf{k} \neq \mathbf{k}^J$.[18] A deviation by firm h from the jointly efficient outcome to some k_h then triggers a discrete fall in the level of investment protection from $\theta^{JC} \geq \theta^J$ to $\Theta(k_h, \mathbf{k}_{-h}^J) < \theta^J$. This will not be profitable to any investor if (7) is fulfilled, since the right–hand side of (7) represents firm h's maximal expected profit of a deviation from k_h^J if there is no investment protection. This mechanism illustrates the general point that *anticipated changes in investment protection* $\Theta^C(\mathbf{k})$ can incentivize strategic investors to behave efficiently.

Surplus division The two incentive compatibility constraints—$\theta^{JC} \geq \theta^J$ for regulation and (7) for investment—establish *lower bounds* on investment protection, so any value of θ^{JC} that exceeds the larger of these two bounds will implement the jointly efficient outcome. But the choice of θ^{JC}

[18] Aisbett, Karp and McAusland (2010a) mention this type of compensation mechanism in passing.

still matters, since it affects the distribution of the surplus between the parties. Under a carve-out compensation scheme $\mathbf{T}^C$ that implements $(\mathbf{k}^J, \theta^J)$, the host country equilibrium expected utility is

$$\tilde{V}(\mathbf{T}^C) \equiv \int_{\underline{\theta}}^{\theta^J} V(\mathbf{k}^J, \theta) dF(\theta) - [F(\theta^{JC}) - F(\theta^J)]\Pi(\mathbf{k}^J), \tag{8}$$

and the equilibrium expected industry profit equals

$$\tilde{\Pi}(\mathbf{T}^C) = F(\theta^{JC})\Pi(\mathbf{k}^J) - R(\mathbf{k}^J). \tag{9}$$

These expressions show in particular that a higher θ^{JC} transfers expected surplus from Home to Foreign by increasing expected compensation payments.

By substituting (8) and (9) into the Nash Product (6), and maximizing over θ^{JC}, we obtain the Nash Bargaining Solution with an efficient market outcome $(\mathbf{k}^J, \theta^J)$ as:

$$F(\theta^{JC}) = \frac{R(\mathbf{k}^J) + \pi^0}{\Pi(\mathbf{k}^J)} + (1-\alpha)\frac{\omega^J - \omega^0}{\Pi(\mathbf{k}^J)}, \tag{10}$$

where ω^J is the expected joint surplus (2) evaluated at $\mathbf{k}^J$, and $\omega^0 \equiv v^0 + \pi^0$ is the expected joint surplus absent an agreement. The negotiated carve-out is larger—that is, θ^{JC} is lower—when Home has more bargaining power (α is larger), operating profits $\Pi(\mathbf{k}^J)$ are larger, and investment costs $R(\mathbf{k}^J)$ are smaller.

4.2.2 Implications for the interpretation of core features of actual agreements

Proposition 1 shows that an appropriately designed agreement with key features that are compatible with actual compensation schemes, not only benefits both parties, but might fully solve the externality problems that these agreements are meant to address. It thus gives an economic foundation for core features of actual agreements.

The role of full compensation As mentioned above, the requirement for any compensation to fully cover foregone operating profits is closely in line with the dictum in international law regarding the state responsibility to "wipe out all the consequences" of compensable acts. Proposition 1 shows that this feature is desirable also from an economic perspective. The simple, but fundamental reason is that full compensation induces the host country to *fully internalize* the consequences for investors of its regulatory decisions. Full compensation for foregone operating profits thus reflects basic principles in both international law and economics.

The exemptions for "legitimate" policy interventions The negotiated Pareto optimal agreement derived above includes exceptions from the compensation requirement for sufficiently severe regulatory shocks. This is closely in line with recent actual agreements that typically contain explicit exceptions for policies "designed and applied" to protect "legitimate" policy objectives. In our setting, it is natural to interpret a policy intervention to be "legitimate" if the regulatory shock is more severe than $\Theta^C(\mathbf{k})$. The legitimacy of the regulation thus depends on the details of the factual situation, in that the carve-out depends on the magnitude of both investment $\mathbf{k}$ and the regulatory shock θ. This also seems to be consistent with actual practice. For instance, Annex B of the *2012 U.S. Model Bilateral Investment Treaty* states that "[t]he determination of whether an action or series of actions by a Party, in a specific fact situation, constitutes an indirect expropriation, requires a case-by-case, fact-based inquiry...". As inquiries necessarily occur after intervention has taken place, it is likely that assessments will depend both on the magnitude of actual investment, and on the severity of the regulatory problem.

Compensation payments A natural way to encourage investment would be to subsidize it, perhaps additionally to protecting investment against regulatory expropriation. Subsidization of foreign investment occurs in practice through general subsidy schemes, and in government contracts with specific firms, but is not part of state-to-state investment agreements. One plausible reason is that the broad scope and long-run nature of investment agreements effectively render it impossible to include subsidization. But the agreements still provide an *implicit* form of investment support in the form of equilibrium compensation payments for $\theta \in (\theta^J, \theta^{JC}]$. Such payments are often seen either as evidence of deliberate violations by the host country of the spirit of the agreement, or as indications of a flawed legal regime that allows investors to extract protection rents. The present framework suggests a very different role: Compensation payments serve not only to prevent over-regulation, but also to stimulate investment, the benefits of which materialize for realizations of θ *other* than those for which the compensation payments are made:

Corollary 1 *Equilibrium compensation payments are required in order to implement the jointly efficient market outcome, and also serve as implicit side-payments in the negotiation of the investment agreement.*

Put differently, when carve-out compensation implements the efficient outcome, there is no need for ex ante subsidy commitments.

Arbitration Under a general compensation scheme, firm-level compensation $T^h(\mathbf{k}, \theta)$ would depend on the investments $\mathbf{k}$ of all firms affected by the regulation and the magnitude θ of the shock. Under a carve-out mechanism, firm-level compensation boils down to estimating foregone operating profit. Under $\mathbf{T}^C$ and efficient investment $\mathbf{k}^J$, the arbitration court must instead decide whether regulation represented a legitimate policy intervention, in the sense of $\theta > \theta^{JC}$, and therefore no compensation should be paid out, or if intervention amounted to regulatory expropriation, in the sense of $\theta \leq \theta^{JC}$, in which case investors should be compensated by their foregone operation profits. Our analysis assumes perfect enforcement, meaning that the arbitration court observes whether $\theta - \theta^{JC}$ is positive or negative. However, fundamental results of our paper do not depend crucially on such perfect enforcement.

Let us introduce a noisy signal by assigning a probability $Q(|\theta - \theta^{JC}|)$ to the event that the court correctly asserts the sign of $\theta - \theta^{JC}$, where Q is strictly increasing in $|\theta - \theta^{JC}|$ for $Q < 1$. By this assumption, the court is more likely to make a correct judgement if θ differs a lot from θ^{JC}. Let Q be a measure of the *quality* of the arbitration court. Given efficient investment $\mathbf{k}^J$, the host country will regulate for all $\theta > \theta^J$, no matter the properties of Q. It will allow production for all $\theta \leq \theta^J$ if and only if

$$V(\mathbf{k}^J, \theta^J) + Q(\theta^{JC} - \theta^J)\Pi(\mathbf{k}^J) \geq 0 \Leftrightarrow (1 - Q(\theta^{JC} - \theta^J))\Pi(\mathbf{k}^J) \leq 0.$$

Hence, implementation of $(\mathbf{k}^J, \theta^J)$ under carve-out compensation requires $Q(\theta^{JC} - \theta^J) = 1$.[19] But efficiency does not dictate perfect enforcement in all states of the world, only that the court is of sufficiently high quality that it can identify overregulation (regulation that occurs for shocks $\theta \leq \theta^J$). Such identification is easier when investment protection θ^{JC} is more extensive by the assumptions on Q.[20] We summarize these results as:

Remark 1 *A carve-out compensation scheme can implement the jointly efficient market outcome also under imperfect enforcement if the quality of the arbitration court is sufficiently high. Implementation is easier if investment protection is more extensive.*

Suppose that the task of the arbitration court is instead to determine whether an intervention was ex post efficient, meaning that the quality of the signal is $Q(|\theta - \theta^J|)$. Then it is impossible to implement the jointly efficient solution unless enforcement is perfect in all states of the world, so that $Q(0) = 1$. Otherwise, $V(\mathbf{k}^J, \theta^J) + Q(\theta^J - \theta^J)\Pi(\mathbf{k}^J) < 0$. On the basis of this observation, the proper task of an arbitration court in our context is to assess whether a policy intervention constituted a violation of the terms of the investment agreement, i.e. whether $\theta < \theta^{JC}$, rather than to assess whether the intervention was unjustifiable on economic grounds, i.e. whether $\theta < \theta^J$.

On the plausibility of the underlying mechanisms Proposition 1 presumes that the agreement is designed to exploit investors' awareness of how their investments affect expected compensation payments. This requires that negotiators understand the functioning of the economy and are able to identify the efficient outcome, which certainly requires sophisticated negotiators. But this assumption is not different from other models with rational economic agents. For instance, it is routinely—but implicitly—assumed in the literature on trade agreements that negotiators understand the complete working of the global general equilibrium system when they draft agreements covering many thousands of tariff bindings. A second behavioral assumption is that investors act

[19] Punitive compensation $\sum_{h=1}^{H} T^h(\mathbf{k}^J, \theta) > \Pi(\mathbf{k}^J)$ can restore regulation incentives if $Q(\theta^{JC} - \theta^J) < 1$. See the discussion of efficient implementation below.

[20] The host country can be more inclined to accepting extensive investment protection under imperfect enforcement relative to the case of perfect enforcement because the marginal increase in compensation payments associated with an increase in investment protection can be smaller in the former than the latter case.

strategically in the sense that they take into account how their investment affects investment protection $\Theta^C(\mathbf{k})$. Whether investors in practice account for such effects is likely to depend on the size of the investment and the structure of the industry. We have also considered the polar case where investors treat investment protection as exogenous to the own investment. Then the agreement can still achieve efficient regulation and investment under robust circumstances, but it is generally impossible to simultaneously implement the Nash Bargaining Solution (10).[21] Implementation of the jointly efficient outcome would then require that the two countries maximize their joint expected surplus.

4.3 When carve-out compensation cannot implement the efficient outcome

The proof of Proposition 1 identifies circumstances under which negotiations over carve-out compensation schemes will implement the jointly efficient outcome. It thus also implicitly identifies two general circumstances under which carve-out compensation mechanisms cannot be expected to implement this outcome. First, the efficient outcome need not be incentive compatible under $\mathbf{T}^C$. This problem arises if deviating from the efficient outcome is sufficiently profitable for at least one firm h that (7) is violated for all $\theta^{JC} \in [\theta^J, \bar{\theta}]$. Second, the bargaining strength of the two negotiating parties can be such that the Nash Bargaining Solution prevents them from achieving ($\mathbf{k}^J$,θ^J) even if it would be technically feasible to do so, i.e., even if (7) is satisfied.

If Nash Bargaining with carve-out compensation does not reach full efficiency, questions arise regarding whether there exist efficient non-carve out mechanisms, how they then differ from carve-out compensation, and whether countries could agree on such alternative mechanisms. To address these questions, we introduce the following intermediate form of compensation scheme:

Definition 3 *A **non-punitive compensation scheme** is a vector of transfers* $\mathbf{T}^N = (T^{1N}, .., T^{HN})$*, where* $T^{hN}(\mathbf{k}, \theta) = \beta(\mathbf{k}, \theta)\Pi^h(\mathbf{k})$ *for all* h *and where* $\beta(\mathbf{k}, \theta) \in [0, 1]$*, to be paid if and only if the host country regulates, without third-party involvement.*

Compensation scheme $\mathbf{T}^N$ is a special case of the general compensation scheme $\mathbf{T}$ by tying compensation payments to operating profit and placing an upper limit—full foregone operating profits—on such compensation. But it is more general than the carve-out scheme $\mathbf{T}^C$ by allowing for intermediary levels of compensation, and by not relying on a threshold for when regulation is compensable. The following Proposition establishes that a carve-out compensation scheme is weakly preferred to a non-punitive scheme even when the carve-out scheme does not implement the jointly efficient outcome (the proof is in Appendix A.2):

Proposition 2 *For any agreement with non-punitive compensation* $\mathbf{T}^N$*, there exists an agreement with carve-out compensation* $\mathbf{T}^C$ *that gives all firms the same expected profit as with* $\mathbf{T}^N$ *(*$\tilde{\Pi}^h(\mathbf{T}^C) = \tilde{\Pi}^h(\mathbf{T}^N)$ *for all* h*), and that gives the host country weakly higher expected utility (*$\tilde{V}(\mathbf{T}^C) \geq \tilde{V}(\mathbf{T}^N)$*).*

[21] The formal statement and proof are omitted for the sake of brevity, but are available on request.

After the realization of the regulatory shock θ, it is efficient to permit production for mild shocks and to regulate for severe shocks (by $V_\theta < 0$). Carve-out compensation $\mathbf{T}^C$ implements such a threshold for regulation by construction, whereas regulation can occur for a non-convex set $M^r(\mathbf{k})$ of shocks under non-punitive compensation $\mathbf{T}^N$. We show in Appendix A.2 how investment protection that yields full compensation for a narrow range of shocks, can be designed to provide investors with the same incentives to invest and the same expected investment profit, as a scheme that awards compensation for a share $\beta(\mathbf{k}, \theta)$ of foregone operating profit for a broader range of shocks. As a consequence, production will be allowed more often for mild shocks, and regulation will occur more frequently for severe shocks. This increase in regulatory efficiency benefits the host country.

Proposition 2 has an important implication: It shows that when the negotiations over a carve-out compensation scheme fails to implement the jointly efficient outcome, it does not help to allow for the significantly more general non-punitive compensation scheme—*it is necessary to introduce features that typically are not found in actual agreements*. We briefly point to several such possibilities.

Negative compensation The proof of Proposition 1 establishes that firm h earns at least the right-hand side of inequality (7) by deviating from k_h^J under any compensation scheme that requires non-negative compensation $T^h(\mathbf{k}, \theta) \geq 0$. Setting $T^h(\mathbf{k}, \theta) < 0$ for $\mathbf{k} \neq \mathbf{k}^J$—that is, requesting investors to pay compensation to the host country if failing to invest the efficient amount—would then facilitate implementation of the efficient outcome by reducing the deviation profit. But such *negative compensation* has a number of unappealing properties. It can reinforce host country incentives to overregulate (out of equilibrium). It can also be difficult to enforce if the regulation has erased the value of investors' assets. As a result, it might be infeasible for the host country to extract compensation payments from investors subsequent to regulation. Furthermore, it might be difficult for the source country government to enforce such an agreement. Because of the implausibility of negative compensation, we maintain the assumption that all compensation is non-negative.

Punitive compensation Another possibility would arise if the agreement allowed for *punitive compensation payments*, that is, payments that exceed foregone operating profits, $T^h(\mathbf{k}^J, \theta) > \Pi^h(\mathbf{k}^J)$. Such compensation would increase the profitability of investing k_h^J and thus increase the left-hand side of (7) relative to what is possible under carve-out compensation. Punitive compensation is not unproblematic, either. First, it yields underregulation if the host country does not regulate for severe shocks $\theta > \theta^J$, for the fear of compensation payments. Second, punitive compensation implies that firms earn more subsequent to regulation than under normal operations. This asymmetry could be the source of a moral hazard problem by which firms invest in assets with high regulatory risk with the sole aim of being regulated and thereby receive punitive compensation payments.[22] We do

[22] Assume that either $\theta = \underline{\theta}$, meaning there is no shock and no regulation, or there is a severe shock $\theta = \bar{\theta} > 0$ with regulation. Let there be one firm in the industry, and assume that this firm has the choice between a safe technology with zero probability of regulation, or a risky technology that yields the severe shock with probability $\zeta > 0$. Holding investment k constant across the two technologies, the net expected benefit of the risky technology over the safe one

not consider firms' endogenous choice of technology in our model by our assumption that $F(\theta)$ is independent of $\mathbf{k}$. We will therefore consider the implications of punitive compensation payments, bearing in mind the possibility of underregulation.

Firm-specific compensation A third possibility could be to give firms *different fractions* of their foregone operating profits. Firm-specific θ^{hJC} could reduce total compensation payments relative to an agreement with uniform θ^{JC}, while still maintaining incentive compatibility of investments. This modification increases host country expected utility of entering into an agreement with efficient implementation $(\mathbf{k}^J, \theta^J)$.

Compensation based on other variables than foregone operating profits In our model, pay-off relevant variables that govern investment and regulation include investment costs $R(\mathbf{k})$ and host country utility $V(\mathbf{k}, \theta)$. Tying compensation to such variables can potentially increase efficiency, as we demonstrate below.

Investment agreements with general compensation schemes Investment agreements that build on general compensation schemes allow punitive compensation, firm-specific compensation payments, and more broad-based compensation than foregone operating profit. For such agreements (the proof is in Appendix A.3):

Proposition 3 *For any agreement with general compensation $\mathbf{T}$, there exists an alternative agreement with general compensation $\hat{\mathbf{T}}$ that yields a threshold for regulation $\hat{\Theta}(\mathbf{k}) \in [\Theta(\mathbf{k}), \Theta^J(\mathbf{k})]$, offers firms the same expected investment profit as in the initial agreement ($\tilde{\Pi}^h(\hat{\mathbf{T}}) = \tilde{\Pi}^h(\mathbf{T})$ for all h), and gives the host country weakly higher expected utility ($\tilde{V}(\hat{\mathbf{T}}) \geq \tilde{V}(\mathbf{T})$).*

That is, Pareto optimal general compensation schemes feature a *threshold for regulation* $\hat{\Theta}(\mathbf{k})$. We show in the proof of the Proposition how one can reshuffle payments in any initial compensation scheme $\mathbf{T}$ such that the host country allows production for all shocks $\theta \leq \hat{\Theta}(\mathbf{k})$, but regulates otherwise, without affecting either incentives to invest or expected investment profits. The host country benefits from the resulting increase in regulatory efficiency. The intuition is analogous to the reasoning above regarding punitive compensation. Moreover, $\hat{\Theta}(\mathbf{k})$ either implies ex post efficient regulation or overregulation. Underregulation can occur only if the agreement stipulates compensation in excess of foregone operating profits for some $\theta > \Theta^J(\mathbf{k})$, which implies that total compensation payments would have to exceed aggregate operating profit $\Pi(\mathbf{k})$ in those events.[23] The profit of firm h is $\Pi^h(\mathbf{k})$ for such realizations of θ, since production is allowed due to underregulation. Reducing compensation down to $\Pi^h(\mathbf{k})$ for all firms for these values of θ would

equals $\zeta(T(k) - \Pi(k))$, which is strictly positive if $T(k) > \Pi(k)$. Given that the investment cost of the risky technology probably is smaller than that of the safe one, the only way the host country can implement a safe technology is to set $T(k) \leq \Pi(k)$ in this case.

[23] Otherwise, $V(\mathbf{k}, \theta) + \sum_{h=1}^{H} T^h(\mathbf{k}, \theta) \leq V(\mathbf{k}, \theta) + \Pi(\mathbf{k}) < V(\mathbf{k}, \Theta^J(\mathbf{k})) + \Pi(\mathbf{k}) = 0$ for all $\theta > \Theta^J(\mathbf{k})$, in which case it is ex post optimal to regulate.

instead induce the host country to regulate. But investment incentives would remain unaffected as every firm would still receive its operating profit for those shocks–albeit now as compensation for regulation. The modification of the compensation scheme thus increases regulatory efficiency by eliminating underregulation without influencing investments or profits, and therefore represents a Pareto improvement.

Proposition 3 derives a robust property of regulation for general compensation schemes. But the implied threshold for regulation can easily be replicated through carve-out compensation. For $\hat{\Theta}(\mathbf{k}) = \Theta^J(\mathbf{k})$, set $\Theta^C(\mathbf{k}) \geq \Theta^J(\mathbf{k})$, and for $\hat{\Theta}(\mathbf{k}) < \Theta^J(\mathbf{k})$, set $\Theta^C(\mathbf{k}) = \hat{\Theta}(\mathbf{k})$. Hence, the inefficiency of carve-out compensation $\mathbf{T}^C$ relative to general compensation $\mathbf{T}$ does not concern incentives to regulate, but instead investment incentives.

To see how a more general compensation scheme than $\mathbf{T}^C$ can improve investment incentives, consider implementation of the efficient outcome $(\mathbf{k}^J, \theta^J)$ under general compensation scheme $\hat{\mathbf{T}}$ with a threshold for regulation $\hat{\Theta}(\mathbf{k})$. The problem of underregulation yields an upper bound $-V(\mathbf{k}^J, \theta)$ to total compensation payments $\sum_{h=1}^{H} \hat{T}'^h(\mathbf{k}^J, \theta)$ under $\hat{\mathbf{T}}$. Three implications follow. First, general compensation payments facilitate implementation of efficient investment by increasing the magnitude of compensation payments relative to carve-out compensation: $-V(\mathbf{k}^J, \theta) > \Pi(\mathbf{k}^J)$ for all $\theta > \theta^J$. Hence, the agreement relies on punitive compensation. Second, compensation is based on the host country's benefit of regulation $-V(\mathbf{k}^J, \theta)$ rather than on foregone operating profits $\Pi(\mathbf{k}^J)$. Finally, the increase in expected compensation payments reduces the host country's net benefit of entering into an agreement. This means that it could be feasible to negotiate an agreement with carve-out compensation and inefficient investment and regulation, but infeasible to negotiate an agreement with general compensation that is efficient.

Remark 2 *Any inefficiency of carve-out compensation relates to an inability to stimulate investment, but not to correct distorted regulation incentives. An increase in efficiency can be achieved, for instance, by tying compensation to the host country benefit of regulation.*

Aisbett, Karp and McAusland (2010a) consider a model in which investment incentives are efficient, but the host country has excessive incentives to regulate. Enforcement of the agreement is imperfect because the arbitration court receives a noisy signal $\tilde{\theta}$ about the regulatory shock θ. The main mechanism is qualitatively similar to (3), except $T^{hC}(\mathbf{k},\tilde{\theta}) = \beta(\mathbf{k})\Pi^h(\mathbf{k})$ for $\tilde{\theta} \leq \Theta^C(\mathbf{k})$, and zero otherwise. The efficient mechanism entails punitive damage payments: $\beta(\mathbf{k}^J) > 1$. Aisbett, Karp and McAusland (2010a) also derive an alternative solution. In this case, regulation always features compensation as a linear combination of operating profits $\Pi^h(\mathbf{k})$ and investment costs $R^h(\mathbf{k})$. The mechanism is efficient because it has two instruments that can be used to incentivize investment and regulation. The efficient scheme still involves excessive compensation. Stähler (2018) derives an alternative mechanism that can implement the globally efficient solution without any information about the magnitude of the shock θ. Ex post efficient regulation is ensured by requiring that the host country always pays compensation $\Pi(\mathbf{k})$. Efficient investment is achieved by paying each firm a compensation payment that causes it to fully internalize all externalities of its decision. Compensation is not generally budget-balanced in this scheme, so Stähler (2018) assumes that an arbitrator evens out the difference.

4.4 Two-way investment flows

We have so far considered a setting in which investment flows in one direction only. But actual investment agreements are reciprocal, and are increasingly often entered into by countries with the potential of serving both as host and source countries for foreign investment. Similarly, older agreements that effectively applied to investment from a developed to a developing country, are becoming increasingly reciprocal due to the economic development of the latter partner. The existence of two-way investment flows has important consequences, both by increasing the scope for investment agreements, and by aligning the interests of the parties to agreements.

Assume that Home and Foreign serve as both hosts and sources of foreign investment, but the two economies are not linked in any other way than through the agreement. Specifically, the profits from investment abroad are unrelated to activities in the domestic economy. Negotiation over an agreement now concerns a pair of general compensation mechanisms $(\mathbf{T}, \mathbf{T}^*)$, where we indicate variables pertaining to Foreign by an asterisk (*). We assume again that the outcome of the negotiations is given by the Nash Bargaining Solution. The negotiated agreement thus maximizes the Nash Product:

$$\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*) \equiv [\tilde{V}(\mathbf{T}) + \tilde{\Pi}^*(\mathbf{T}^*) - v^0 - \pi^{*0}]^{\alpha}[\tilde{V}^*(\mathbf{T}^*) + \tilde{\Pi}(\mathbf{T}) - v^{*0} - \pi^0]^{1-\alpha}.$$

An alternative would be to negotiate two separate agreements over one-way investment flows. But since the countries can replicate any pair of agreements that would result from such negotiations, the agreement covering two-way investment flows is obviously at least as efficient. Yet, by the assumed separability of the two economies, a two-way agreement does not allow for the internalization of any additional externalities. Consequently, the same conditions for efficiency, $\theta^{JC} \geq \theta^J$ and (7), still apply in both countries under two-way investment. Even so, the agreement covering two-way investment flows is typically *more* efficient than a pair of agreements over one-directional flows (the proof is in Appendix A.4):

Proposition 4 *Concerning investment agreements with two-way investment flows:*
(1) A carve-out scheme $(\mathbf{T}^C, \mathbf{T}^{*C})$ *implements the jointly efficient market outcome* $(\mathbf{k}^J, \theta^J, \mathbf{k}^{*J}, \theta^{*J})$ *and maximizes the Nash Product* $\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*)$ *in a broader set of circumstances than with two separate agreements that each covers one-way investment flows.*
(2) A carve-out scheme with symmetric investment protection ($\theta^{JC} = \theta^{*JC}$) *can implement the Nash Bargaining Solution even if countries are asymmetric, in a robust set of circumstances.*

To see why agreements with two-way flows tend to be strictly more efficient, observe that with two-way flows, the surplus of Home equals

$$\tilde{V}(\mathbf{T}^C) + \tilde{\Pi}^*(\mathbf{T}^{*C}) = \int_{\underline{\theta}}^{\theta^J} V(\mathbf{k}^J, \theta) dF(\theta) - [F(\theta^{JC}) - F(\theta^J)]\Pi(\mathbf{k}^J) + F^*(\theta^{*JC})\Pi^*(\mathbf{k}^{*J}) - R^*(\mathbf{k}^{*J})$$

under an agreement with carve-out compensation that implements the efficient outcome. The second term in the above expression represents the expected compensation payments to Foreign investors, and the sum of the last two terms is the expected industry profit from Home firms' outward investments in Foreign. An increase in investment protection θ^{JC} that makes Home worse off by increasing compensation payments can be exactly compensated by a corresponding increase in θ^{*JC} that increases the profitability of foreign direct investment, to keep Home and Foreign equally well off as before. It does not matter to the two countries if they agree on high levels of investment protection in both countries or low levels, as long as those changes do not affect investment and regulation, since what each country gains on its inward investment is lost on its outward investment, or vice versa. Incentives regarding investment protection are therefore aligned under two-way investment flows, which was not the case under one-way flows. The alignment of preferences implies that countries can achieve an agreeable distribution of surplus in a *broader* set of circumstances than was feasible under one-way investment flows.

Turning to Part (2) of Proposition 4, the jointly optimal outcomes $(\mathbf{k}^J, \theta^J)$ and $(\mathbf{k}^{*J}, \theta^{*J})$ will normally differ across countries because of asymmetries. Yet, countries almost invariably negotiate agreements that apply symmetrically to both countries. The second part of Proposition 4 shows that such *contractual symmetry does not necessarily imply contractual inefficiency even if countries are asymmetric*. Formally, the reason is that the division of surplus is constant for a combination of $(\theta^{JC}, \theta^{*JC})$. Therefore, $\theta^{JC} = \theta^{*JC}$ maximizes $\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*)$ under certain conditions, despite cross-country differences.

4.5 The economy-wide coverage of investment agreements

A defining characteristic of actual investment agreements is their economy-wide scope.[24] To see the benefit of such a feature, we now extend our baseline model of one-directional investment to allow for a range of industries $i \in \{1, ..., I\}$, each of which exposes the host country to a separate shock $\theta_i \in [\underline{\theta}_i, \bar{\theta}_i]$ with cumulative distribution $F^i(\theta_i)$ and density $f^i(\theta_i)$. Industry i has H_i Foreign-owned firms that invest $\mathbf{k}_i = (k_{i1}, .., k_{iH})$ in Home. The industries are functionally separable, so that $V^i(\mathbf{k}_i, \theta_i)$ is the host country utility of allowing production in industry i, and $\Pi^i(\mathbf{k}_i) \equiv \sum_{h=1}^{H_i} \Pi^{ih}(\mathbf{k}_i)$ is the associated operating profit in this industry. The separate industries can be thought of as mining, pharmaceuticals, energy generation, and so forth, that can expose the host country to uncorrelated regulatory shocks.

If the carve-out compensation scheme implements $(\mathbf{k}_i^J, \theta_i^J)$ for all industries, then the host coun-

[24] An interesting feature of some recent agreements is that they contain exceptions for certain industries, such as tobacco.

try expected utility is

$$\tilde{V}(\mathbf{T}^C) \equiv \sum_{i=1}^{I}\{\int_{\underline{\theta}_i}^{\theta_i^J} V(\mathbf{k}_i^J,\theta_i)dF^i(\theta_i) - [F^i(\theta_i^{JC}) - F^i(\theta_i^J)]\Pi^i(\mathbf{k}_i^J)\},$$

and the associated expected source country profit becomes

$$\tilde{\Pi}(\mathbf{T}^C) \equiv \sum_{i=1}^{I}\{F^i(\theta_i^{JC})\Pi^i(\mathbf{k}_i^J) - R^i(\mathbf{k}_i^J)\}.$$

The negotiated investment protection solves

$$\sum_{i=1}^{I} F^i(\theta_i^{JC})\Pi^i(\mathbf{k}_i^J) = \sum_{i=1}^{I}(R^i(\mathbf{k}_i^J) + \pi_i^0) + (1-\alpha)\sum_{i=1}^{I}(\omega_i^J - \omega_i^0).$$

Just as in the case with two-way investment flows, investment protection θ_i^{JC} is not uniquely defined. Instead, the host country is willing to trade off more industry protection (fewer carve-outs) in some industries (or shock dimensions) against less investment protection (more carve-outs) in others. This substitutability makes it easier to establish appropriate carve-outs in every industry, and to negotiate an agreeable distribution of surplus. We summarize these findings as:

Proposition 5 *A carve-out compensation scheme implements the jointly efficient market outcome with Nash Bargaining in a broader set of circumstances than with industry-specific agreements. The substitutability of carve-outs across industries/shocks facilitates negotiation of an efficient outcome.*

5 Other standard provisions

The analysis above has sought to capture important features of regulatory expropriation provisions in investment agreements. We now explore basic features of three other salient provisions in the agreements: National Treatment (Section 5.1), compensation for direct expropriation (Section 5.2), and the right for investors to pursue investment disputes against host countries (Section 5.3). To simplify the exposition, we consider the case of one-directional investment, although the results have analogous counterparts for agreements with two-way investment.

5.1 National Treatment

The far-reaching compensation mechanisms in investment agreements have been heavily criticized. Some argue that the *sole* role of such agreements should be to prevent discriminatory treatment of foreign relative to domestic investment. For instance, Stiglitz (2008, p. 249) writes "non-discrimination provisions will provide much of the security that investors need without compromising the ability of democratic governments to conduct their business." In this section, we examine the consequences of a National Treatment (NT) provision that aims to prevent discrimination between foreign and domestically owned firms, and how far such a clause would take the parties toward solving investment and regulatory distortions.

An obvious limitation of non-discrimination clauses is that they only have bite if there exist domestically owned firms that operate under sufficiently "like circumstances" to those facing the foreign investors that have been subject to regulation. We therefore assume that the host country features a domestically-owned industry (indicated by superscript D), in addition to the foreign-owned industry (now indicated by superscript F). The two industries are identical in terms of demand and production structures, suffer from the same country-specific shock, and therefore reasonably produce under "like circumstances" for the purpose of an NT provision. We assume for notational simplicity that the two industries are economically unrelated to one another. The host country fully internalizes the consequences of regulation for the profits of its domestic industry, while it completely disregards the impact on foreign profit of regulatory intervention.

Absent an agreement, the host country would discriminate against the foreign industry in the sense that the foreign industry would be regulated more frequently than the domestic industry for identical investment portfolios $\mathbf{k}$ in both industries: $\Theta^F(\mathbf{k}) = \Theta(\mathbf{k}) < \Theta^J(\mathbf{k}) = \Theta^D(\mathbf{k})$. However, investment will differ between the two industries because of regulatory discrimination, so one cannot say for certain which of the two is regulated more often in equilibrium.

5.1.1 NT only

We represent the NT only clause by an agreement that only contains the requirement that the domestic industry must be regulated whenever there is regulation of the foreign industry. We assume that the clause is binding, so that the threshold for regulation is the same for both industries. Because there are no compensation payments, the equilibrium investment protection under an NT only provision, $\Theta^{NTO}(\mathbf{k}^D, \mathbf{k}^F)$, is implicitly defined by

$$V(\mathbf{k}^F, \Theta^{NTO}(\mathbf{k}^D, \mathbf{k}^F)) + V(\mathbf{k}^D, \Theta^{NTO}(\mathbf{k}^D, \mathbf{k}^F)) + \Pi(\mathbf{k}^D) \equiv 0,$$

where $\mathbf{k}^D$ and $\mathbf{k}^F$ are vectors of arbitrary domestic and foreign investment. The NT provision works very differently from a carve-out policy. From the viewpoint of the incentives to regulate, it is as if the host country *always* pays compensation, and such compensation is paid on the basis of domestic surplus $V(\mathbf{k}^D, \theta) + \Pi(\mathbf{k}^D)$ instead of $\Pi(\mathbf{k}^F)$ as would be the case under carve-out compensation. And because host country utility is negative for all sufficiently serious shocks, it follows that

$$\Theta^F(\mathbf{k}) = \Theta(\mathbf{k}) < \Theta^{NTO}(\mathbf{k}, \mathbf{k}) < \Theta^J(\mathbf{k}) = \Theta^D(\mathbf{k})$$

because investment portfolios are the same in both industries, $\mathbf{k}^D = \mathbf{k}^F = \mathbf{k}$, by symmetry and the fact that they have the same threshold for regulation.

Applied to the present setting, Stiglitz' (2008) assertion that NT provisions are sufficient to improve the protection of foreign investment relative to the case when there is no investment protection, is correct: $\Theta^{NTO}(\mathbf{k}, \mathbf{k}) > \Theta^F(\mathbf{k})$. This can increase foreign investment and is likely to benefit foreign investors.[25]

But an NT provision will generally also affect domestic policies. The host government adapts to the NT clause by reducing investment protection of the domestic industry: $\Theta^{NTO}(\mathbf{k}, \mathbf{k}) < \Theta^D(\mathbf{k})$. The NT only clause has two direct negative consequences for the host country. First, it regulates foreign investors less frequently than what is unilaterally optimal. Second, it induces the host country to regulate the domestic industry more often than it would desire. But the agreement also affects domestic and foreign investment. Under reasonable circumstances, domestic investment falls, and foreign investment increases. If there is underinvestment from the point of view of the host country in both industries, the only possible benefit from an NT only agreement is the increase in foreign investment.

Observation 2 *Relative to the benchmark without an agreement, an agreement comprising only an NT provision is likely to benefit the source country through increased investment protection. But this agreement can have negative consequences for the host country by triggering overregulation of the domestic industry and underregulation of the foreign industry, and by distorting domestic investment.*

5.1.2 Carve-out compensation in relation to NT

An alternative to an NT only agreement is to have only carve-out compensation as in the earlier analysis. Carve-out compensation can replicate the host country incentive to regulate foreign investors and the foreign investment incentives that would arise under an NT only agreement, but without affecting the domestic industry. For the case where the NT only provision increases host country utility associated with the domestic industry relative to the case without an agreement, it is optimal also to include an NT provision in the carve-out agreement. In the opposite case, it is not optimal to include an NT provision. NT reduces the distortion of the regulation of the foreign industry by distorting the regulation of the domestic industry. A carve-out compensation scheme allows the parties to directly correct the former distortion. We collect these observations in the following Proposition:

Proposition 6 *The negotiated agreement will never consist of an NT provision only, as long as carve-out compensation can be included in the agreement.*

[25] This holds unambiguously if there is only one investor.

Proposition 6 is consistent with the observation that investment agreements do not rely on NT provisions alone. The role of NT in investment agreements is very different from that normally played by NT in trade agreements, where NT is seen as an instrument aimed at preventing the contracting nations from opportunistically exploiting contractual incompleteness and thereby undermining bargaining concessions. NT renders such opportunistic behavior less attractive by effectively forcing the importing country to distort also its domestic production if it wants to distort trade.[26] In our setting, there is no scope for opportunistic behavior since the compensation requirements are completely specified and costlessly enforceable. Instead, NT essentially serves to extend the commitment possibilities associated with enforcement to the domestic sector.

Remark 3 *Incorporating an NT provision in an investment agreement allows a host country to indirectly use the enforcement mechanism underlying the agreement to address distorted investment incentives in its domestic sector.*

In the model by Aisbett, Karp and McAusland (2010b), the investor can pay the host country a fee for investment protection under the agreement. This fee is effectively an investor-specific tax on production. The equilibrium features complete investment protection because inefficient dispute settlement implies that there are gains from trade in exchanging more investment protection for a higher tax.[27] Including an NT clause in the agreement cannot be efficient because it limits the possibility for investment-specific taxes and may distort the host country tax system. We do not consider investment-specific taxes in our model. Then, there can be a rationale for including NT clauses as a complement to compensation payments because they can increase efficiency by reducing host country regulatory distortions. Kohler and Stähler (2019) compare an agreement with exogenous investment protection sustained by compensation requirements, to an agreement that solely relies on NT for investment protection, in a model where a tax increase between two periods can trigger compensation payments. They show that NT is better than compensation payments from a total surplus perspective if the domestic industry affected by NT is large relative to the foreign industry. In light of Proposition 6, the efficiency of NT relative to compensation payments is likely to depend on the assumption of exogenous investment protection under the compensation agreement in the Kohler and Stähler (2019) framework.

5.2 Direct expropriation

We have thus far been concerned with host country intervention in the form of regulatory measures that prevent firms from continuing their operations. But investment agreements also contain stipulations regarding *direct* expropriation. These are measures through which the host country

[26] See Horn (2006), Saggi and Sara (2008), and Horn, Maggi and Staiger (2010).

[27] See also Hermalin (1995) for a characterization of an efficient investment-specific tax in the context of direct expropriation.

seizes direct control of the firms' assets. Most agreements do not distinguish between compensation requirements for lawful direct and indirect expropriation. For instance, there should be full compensation for both types of expropriation. However, recent agreements include carve-outs that apply only to indirect expropriation.[28] This observed difference raises the question whether direct and indirect expropriations should be treated differently in investment agreements.

Protecting investors from direct expropriation has intuitive appeal if such measures are viewed as opportunistic take-overs of assets that are not pursued to achieve some reasonable regulatory objective. But direct expropriations are sometimes defended as legitimate exercises of national sovereignty, and claims for compensation are contested.[29] As we will see, the latter argument has some economic merit, in that carve-outs from protection against direct expropriation can be means of reducing inefficient overregulation.

To create an incentive for direct expropriation in our setting, let $\Pi^x(\mathbf{k}) > 0$ be the value to the host country of seizing the industry's assets. But the host country cannot operate these assets more efficiently than investors, so $\Pi^x(\mathbf{k}) \leq \Pi(\mathbf{k})$.[30] The host country still has the option of regulating. Denote by $\Theta^x(\mathbf{k}) \in (\Theta(\mathbf{k}), \Theta^J(\mathbf{k})]$ the threshold for when the host country would expropriate the assets rather than regulate, if it did not have to pay compensation payments for either intervention. Since direct expropriation yields the host country utility $V(\mathbf{k},\theta) + \Pi^x(\mathbf{k})$, and regulation yields 0, $\Theta^x(\mathbf{k})$ is given by

$$V(\mathbf{k},\Theta^x(\mathbf{k})) + \Pi^x(\mathbf{k}) \equiv 0$$

if $V(\mathbf{k},\bar{\theta}) + \Pi^x(\mathbf{k}) \leq 0$, and by $\Theta^x(\mathbf{k}) = \bar{\theta}$ otherwise. Assuming that the host country seizes the assets if and only if doing so is strictly beneficial relative to the other policy options, the host country prefers direct expropriation to non-intervention and regulation for all $\theta < \Theta^x(\mathbf{k})$, but it regulates the industry for all $\theta \geq \Theta^x(\mathbf{k})$. Investors will be expropriated one way or the other absent an agreement, and therefore $\mathbf{k}^0 = \mathbf{0}$.

To incorporate a provision concerning direct expropriation in the agreement, consider an extended carve-out compensation mechanism $\mathbf{T}^C = (\mathbf{T}^{rC}, \mathbf{T}^{xC})$, where $\Theta^{rC}(\mathbf{k})$ is the threshold for when foreign investors are eligible for compensation payments for host country regulation, and $\Theta^{xC}(\mathbf{k})$ is the threshold for compensation payments for direct expropriation. In particular, $\Theta^{xC}(\mathbf{k}) = \bar{\theta}$ for all $\mathbf{k}$ under complete investment protection from direct expropriation. Seizing the industry's assets is always a (weakly) dominated strategy relative to non-intervention in that case:

$$V(\mathbf{k},\theta) + \Pi^x(\mathbf{k}) - \Pi(\mathbf{k}) \leq V(\mathbf{k},\theta) \text{ for all } \theta.$$

[28] For instance, while Annex 8-A CETA states that "non-discriminatory measures...that are designed and applied to protect legitimate public objectives...do not constitute indirect expropriations," there is no corresponding restriction on what constitutes direct expropriation.

[29] Direct expropriations were particularly common during the 1950s and 1960s, but see Hajzler (2012) for more recent examples.

[30] The case of $\Pi^x(\mathbf{k}) > \Pi(\mathbf{k})$ seems implausible. But note that overregulation would not occur in this case.

Hence, *complete investment protection is effectively a ban on direct expropriation.*

To see how a carve-out for direct expropriation can improve efficiency, assume that there is a range of θ for which the host country will directly expropriate as long as this is not compensable, but for which it can also regulate without compensation, that is, $\Theta^{rC}(\mathbf{k}) < \Theta^{x}(\mathbf{k})$. If $\Theta^{xC}(\mathbf{k})$ now is reduced from $\bar{\theta}$ to $\Theta^{rC}(\mathbf{k})$, it becomes optimal for the host country to seize the assets instead of regulating the industry for all regulatory shocks $\theta \in (\Theta^{rC}(\mathbf{k}), \Theta^{x}(\mathbf{k}))$, since it does not have to compensate investors either way, and it prefers expropriation to regulation in this range. The host country surplus was 0 before the change because of regulation, but now increases to $V(\mathbf{k},\theta) + \Pi^{x}(\mathbf{k}) > 0$ because of direct expropriation. The change does not affect investors because they have the same effective investment protection as before. The policy change therefore represents a Pareto improvement.

We now establish thresholds for when complete investment protection is optimal, and when it would be more efficient to allow some uncompensated direct expropriation in an agreement. Denote by $\hat{\mathbf{k}}$ the equilibrium investment portfolio under agreement $\mathbf{T}^{C} = (\mathbf{T}^{rC}, \mathbf{T}^{xC})$. Then $\hat{\theta}^{rC} \equiv \Theta^{rC}(\hat{\mathbf{k}})$ characterizes the equilibrium investment protection from regulatory expropriation under $\mathbf{T}^{C}$, and $\hat{\theta}^{x} \equiv \Theta^{x}(\hat{\mathbf{k}})$ is the corresponding threshold below which it is ex post optimal for the host country to seize the assets if it can do so without paying compensation. Observe that $\hat{\theta}^{x}$ is lower, the lower is the value $\Pi^{x}(\hat{\mathbf{k}})$ of the expropriated assets. Recall that $\hat{\theta}^{J} \equiv \Theta^{J}(\hat{\mathbf{k}}) \geq \hat{\theta}^{x}$ is the ex post efficient threshold for regulation given the equilibrium investment portfolio. The proof of the following result can be found in Appendix A.5:

Proposition 7 *Concerning compensation for direct expropriation in investment agreements:*
(1) Complete protection against direct expropriation ($\Theta^{xC}(\mathbf{k}) = \bar{\theta}$ for all $\mathbf{k}$) is Pareto optimal under carve-out scheme $\mathbf{T}^{C} = (\mathbf{T}^{rC}, \mathbf{T}^{xC})$ if either:
(i) The scheme features extensive protection from regulatory expropriation ($\hat{\theta}^{rC} \geq \hat{\theta}^{J}$); or
(ii) The value of expropriated assets is small ($\hat{\theta}^{x} \leq \hat{\theta}^{rC} < \hat{\theta}^{J}$).
(2) The Pareto optimal scheme features the same investment protection for direct and indirect expropriation ($\Theta^{xC}(\mathbf{k}) = \Theta^{rC}(\mathbf{k})$ for all $\mathbf{k}$) if the value of expropriated assets is large ($\hat{\theta}^{rC} < \hat{\theta}^{x} \leq \hat{\theta}^{J}$).

The conditions for when complete protection from direct expropriation is optimal are intuitive. If $\hat{\theta}^{J} \leq \hat{\theta}^{rC}$ or $\hat{\theta}^{x} \leq \hat{\theta}^{rC}$, there are two potential consequences of allowing uncompensated direct expropriation: the host country seizes the assets and maintains production either when it would have been better to regulate (for $\theta > \hat{\theta}^{J}$), or when it would have been better not to intervene (for $\theta < \hat{\theta}^{J}$). Direct expropriation *reduces joint surplus* in both cases. The host country could be better off by reducing investment protection, but this does not create enough surplus to compensate the source country. But in an agreement with large carve-outs ($\hat{\theta}^{rC} < \hat{\theta}^{J}$), allowing direct expropriation can be a way to reduce the problem of overregulation if such expropriation preserves enough value ($\hat{\theta}^{x} >$

$\hat{\theta}^{rC}$). Proposition 7 shows how to optimally modify the agreement to allow for direct expropriation, namely establish the *same* carve-outs for indirect and direct expropriation.

While carve-outs for direct expropriation can reduce overregulation, they can also have undesirable side effects that might explain the aversion to such carve-outs. The first issue is an implication of the fact that direct expropriation is a weakly dominated strategy whenever it causes compensation payments, because those payments exceed the value of the seized assets. Therefore, there will be no compensation payments for direct expropriation that occurs in equilibrium. The possibility for the host country to seize assets without paying compensation may create an incentive to behave opportunistically that would not arise under complete protection from direct expropriation.[31]

Another troubling aspect of direct expropriation is that government takings generally reduce the value of the seized assets. Therefore, direct expropriation is not the jointly most efficient way of avoiding overregulation in an investment agreement. Consider an initial agreement with equilibrium investment $\hat{\mathbf{k}}$ that features equilibrium expropriation. By the proof of Proposition 7, any such agreement has the same equilibrium investment protection $\hat{\theta}^C < \hat{\theta}^x$ for both types of expropriation. Suppose now that the agreement was modified to entail full protection from direct expropriation and protection $\hat{\theta}^x$ from indirect expropriation.[32] Now there will never be any direct expropriation, and the firm continues its production for all $\theta \in (\hat{\theta}^C, \hat{\theta}^x]$ instead of being expropriated. Joint efficiency increases if $\Pi(\hat{\mathbf{k}}) > \Pi^x(\hat{\mathbf{k}})$. Alas, the modification does not constitute a Pareto improvement. Instead, the host country loses $[F(\hat{\theta}^x) - F(\hat{\theta}^C)]\Pi^x(\hat{\mathbf{k}})$ from not being able to seize assets anymore, whereas foreign investors gain the larger amount $[F(\hat{\theta}^x) - F(\hat{\theta}^C)]\Pi(\hat{\mathbf{k}})$ from improved investment protection.

Corollary 2 *Any agreement with carve-out compensation* $\mathbf{T}^C$ *that features direct expropriation in equilibrium is jointly inefficient, but can nevertheless be Pareto optimal.*

The reason why the two countries in our setting would negotiate a relatively less efficient agreement with direct expropriation over a more efficient agreement with higher investment protection would

[31] Suppose that an arbitration court cannot perfectly observe whether host country intervention was legitimate ($\theta \geq \hat{\theta}^C = \hat{\theta}^{xC} = \hat{\theta}^{rC}$) or illegitimate ($\theta < \hat{\theta}^C$), similar to the analysis in Section 4.2. Let $Q \in (0, 1]$ be the exogenous probability of a correct assessment by the court. Consider a shock that is sufficiently weak that it would be efficient to allow production, but such that the host country would intervene either by regulation or by direct expropriation if it could do so without paying compensation: $\theta \in (\Theta(\hat{\mathbf{k}}), \hat{\theta}^C)$. The host country value of allowing production is $V(\hat{\mathbf{k}},\theta) < 0$. The expected cost of shutting down the industry is $Q\Pi(\hat{\mathbf{k}})$. The host country then correctly prefers to allow production rather than to regulate if the quality of the court is sufficiently high in the sense that $Q \geq Q^r \equiv \frac{-V(\hat{\mathbf{k}},\hat{\theta}^C)}{\Pi(\hat{\mathbf{k}})}$. Instead, the host country expected surplus is $V(\hat{\mathbf{k}},\theta) + \Pi^x(\hat{\mathbf{k}}) - Q\Pi(\hat{\mathbf{k}})$ if it seizes control of the industry's assets, which yields a threshold $Q \geq Q^x \equiv \frac{\Pi^x(\hat{\mathbf{k}})}{\Pi(\hat{\mathbf{k}})} \geq Q^r$ for when it is better to correctly allow production than to expropriate. Because direct expropriation can be a more attractive policy option than regulation, it requires more of the arbitration court to prevent opportunistic host country behavior than in an agreement with complete protection against direct expropriation. The difference $Q^x - Q^r$ in quality requirements is larger when carve-outs are larger: Q^r is smaller when $\hat{\theta}^C$ is smaller, but Q^x is independent of $\hat{\theta}^C$. We leave the issue of incompletely informed arbitration courts to further research, and refer instead to Ossa, Staiger and Sykes (2019) for a comprehensive analysis in the context of direct expropriation.

[32] It is easy to verify that $\hat{\mathbf{k}}$ can be sustained as an equilibrium with these higher levels of investment protection.

be surplus distribution. Carve-outs from direct expropriation is a means by which a host country with bargaining power can increase its share of the total pie.

5.3 Investor-state versus state-state dispute settlement

A highly contentious feature of investment agreements is that they typically allow foreign *investors* to file arbitration requests against host countries (ISDS). This is an unusual feature for international treaties, which normally only include state-state dispute settlement (SSDS). A common claim is that ISDS induces investors to pursue too many disputes, and it would therefore be in the interest of host countries to abandon ISDS in their agreements. Casual empiricism certainly suggests that there might be something to this claim, in that virtually all of the over 940 known investment disputes have been initiated by investors, despite the fact all investment agreements allow for SSDS.[33] Some countries have reduced the possibility for ISDS by making the use of this instrument more difficult or time consuming. Others have completely removed the ISDS option in their agreements. For instance, Canada did this in the recently renegotiated NAFTA.

A common explanation for why investment agreements came to allow legal standing for private foreign investors is that disputes between states give rise to political and/or diplomatic frictions that do not arise when private parties request arbitration.[34] These frictions are costly as such. They can also reduce the incentives for source countries to enforce host country compensation obligations, they might affect the parties' willingness to enter into agreements, and they might affect the investment protection the agreements provide. In what follows, we will use a very simple extension of the analysis above to examine these effects, where the source country government is exposed to a cost $\mathcal{A} > 0$ for pursuing a dispute.[35] To maintain consistency with the previous notation, we assume that requests for arbitration are coordinated under ISDS if the industry consists of multiple firms, and that arbitration pursued by the source country concerns compensation for all its affected investors under SSDS.

To illuminate the rationale for including ISDS in an agreement, consider an initial agreement with carve-out compensation $\mathbf{T}^C$ that contains SSDS only. Let $\Theta^C(\mathbf{k})$ be the investment protection of the initial agreement, $\hat{\mathbf{k}}$ the equilibrium investment, $\hat{\theta}^C \equiv \Theta^C(\hat{\mathbf{k}})$ the equilibrium investment protection, and let $\hat{\theta}^J \equiv \Theta^J(\hat{\mathbf{k}})$.

Consider first the case where the source country arbitration costs are so large, $\mathcal{A} > \Pi(\hat{\mathbf{k}})$, that it would *not* find it worthwhile to pursue a dispute under SSDS, even if the agreement would stipulate payment of compensation. Since the host country knows that compensation requirements will not

[33] See www.investmentpolicyhub.unctad.org/ISDS for the full list of ISDS cases.

[34] See Vandevelde (1998). Sykes (2005) additionally points to the possibility that some investors may be unable to offer enough political benefits to their governments to induce the governments to pursue disputes. Furthermore, the resolution of state-to-state disputes might not benefit investors, for instance if they come through political settlements, or if the source country government does not pass on compensation payments to investors.

[35] Disputes also give rise to legal process costs that can be substantial. But these arise for both investors and the source country government, and are for simplicity disregarded here.

be credibly enforced, the only equilibrium is $\hat{\mathbf{k}} = \mathbf{k}^0$ with the threshold for regulation $\theta^0 = \Theta(\mathbf{k}^0)$. Here the agreement is meaningless with SSDS only. Allowing for ISDS in the agreement effectively takes us back to the analysis in Section 4. Hence, *ISDS might prevent the complete unraveling of an agreement by increasing enforcement* to the mutual benefit of the host and source country.

Now turn to the case where arbitration costs are sufficiently small, $\mathcal{A} \leq \Pi(\hat{\mathbf{k}})$, that the agreement $\mathbf{T}^C$ is enforced under SSDS. If investment protection is weak in the SSDS agreement, in the sense that $\hat{\theta}^C \leq \hat{\theta}^J$, the credible threat of having to compensate investors induces the host country to allow production if $\theta \leq \hat{\theta}^C$. The arbitration costs do not matter since there are no disputes in equilibrium. An exogenous introduction of ISDS then has no effect. If instead investment protection is sufficiently extensive, $\hat{\theta}^C > \hat{\theta}^J$, there is arbitration and compensation payments in equilibrium for $\theta \in (\hat{\theta}^J, \hat{\theta}^C]$. The host country expected utility of the initial agreement then is

$$\int_{\underline{\theta}}^{\hat{\theta}^J} V(\hat{\mathbf{k}}, \theta) dF(\theta) - [F(\hat{\theta}^C) - F(\hat{\theta}^J)]\Pi(\hat{\mathbf{k}}), \tag{11}$$

whereas the source country expected utility is

$$F(\hat{\theta}^C)\Pi(\hat{\mathbf{k}}) - R(\hat{\mathbf{k}}) - [F(\hat{\theta}^C) - F(\hat{\theta}^J)]\mathcal{A}. \tag{12}$$

The introduction of ISDS into the agreement will not affect arbitration or the host country behavior, all else equal. This implies that also investment incentives and profits remain the same as before. But the source country will still be affected, since with ISDS it can rely on investors to enforce the agreement and thereby *avoid the expected arbitration costs* $[F(\hat{\theta}^C) - F(\hat{\theta}^J)]\mathcal{A}$ it would face with SSDS only. The introduction of ISDS therefore represents a Pareto improvement. Because an agreement with ISDS can replicate any agreement $\mathbf{T}^C$ with SSDS without affecting the host country negatively and with positive effects for the source country, the Nash Bargaining Solution to these negotiations will yield ISDS.

The host country typically is not indifferent with regard to ISDS or SSDS because the dispute settlement system will matter in negotiations over investment protection. In particular, the source country is likely to accept lower investment protection $\hat{\theta}^C$ under SSDS than ISDS to save on arbitration costs. Since the host country expected utility is strictly decreasing in $\hat{\theta}^C$, see (11), *the host country would prefer SSDS if this yields less investment protection.* We summarize these findings as:

Proposition 8 *Concerning dispute settlement in investment agreements:*
(1) Both countries could benefit from including ISDS if SSDS causes complete unraveling of an agreement.
(2) When the agreement does not unravel with SSDS, the source country prefers ISDS, but the host country prefers SSDS if this yields reduced investment protection.
(3) Negotiations over both investment protection and dispute settlement will implement ISDS.

Proposition 8 is consistent with the fact that almost all investment agreements include ISDS. What seems more difficult to explain, is the tendency to exclude ISDS in more recent agreements. An example is Canada's removal of ISDS in the renegotiation of NAFTA. Canada could have had a unilateral interest to exclude ISDS if it expected to be a net recipient of foreign investment from the US and Mexico because SSDS might then reduce the number of investment disputes. Even so, the host country benefits would be insufficient to cover source country costs of SSDS in our model. But NAFTA is not solely about investment protection. A host country may have enough leverage to convince the other treaty countries to accept SSDS in a broader agreement, if it is willing to make concessions in other dimensions. Canada did indeed make such concessions during the renegotiation of NAFTA, for instance regarding trade in dairy products.[36]

Ossa, Staiger and Sykes (2019) investigate dispute settlement in trade and investment agreements more broadly, including the choice between ISDS and SSDS. Subject to a taking, the firm (under ISDS) or its representative government (under SSDS) can request arbitration. The arbitration court then receives a noisy signal whether the host country intervention was efficient (which would correspond to $\theta > \theta^J$ in our model) or inefficient ($\theta \leq \theta^J$). The host country's decision is overturned if regulation is found to be inefficient, but sustained otherwise. The foreign firm has a stronger incentive to request arbitration than the foreign government by an assumption that the perceived benefit of winning is larger for the firm than the government.[37] The firm's stronger incentive to request arbitration will dampen the host country's desire to expropriate the firm under ISDS. But there is also an increased likelihood of non-intervention under ISDS because of the probability of an incorrect court ruling in the complainant's favor. The net effect on efficiency is ambiguous in general, but the authors identify a set of conditions under which the first positive effect of ISDS dominates the second negative effect. This result shows that the efficiency of ISDS does not depend crucially on perfect enforcement.

6 Policy discussion

The analysis above has consequences for core policy issues regarding investment agreements. We discuss regulatory chill and the scope for agreements and their distributional effects.

[36] Of course there could be other reasons for excluding ISDS than the ones we discuss in this paper. Severe criticism has been directed at for instance the possibility for investors to request arbitration outside host countries' legal systems, arbitrators' alleged partiality, the lack of appeal possibilities, the lack of transparency of the arbitration proceedings and outcomes, and the incoherence of the case law.

[37] An assumption that the host country government has an incremental fixed arbitration cost under SSDS is formally equivalent to an assumption that investors and the government face the same arbitration cost, but the government places a lower weight on the foregone operating profits than investors do.

6.1 Regulatory chill

A common concern in the policy debate is that investment agreements cause *regulatory chill.* This concept is rarely precisely defined, but can be given two natural interpretations within the context of our model. *Domestic regulatory chill* will be said to occur if an agreement prevents a host country from undertaking a policy intervention that it would make absent compensation requirements. This seems to capture the sense in which the term typically is applied in the policy debate. By way of Proposition 3, a Pareto optimal agreement indeed causes domestic regulatory chill for all shocks $\theta \in (\Theta(\mathbf{k}), \hat{\Theta}(\mathbf{k})]$, where $\hat{\Theta}(\mathbf{k}) \geq \Theta(\mathbf{k})$ is the threshold for regulation. *Joint regulatory chill* occurs if an agreement induces the host country to allow production in situations where regulation would have been ex post jointly optimal, for shocks $\theta > \Theta^J(\mathbf{k})$. Proposition 3 directly implies:

Corollary 3 *A Pareto optimal agreement implements ex post jointly efficient production for* $\theta \leq \hat{\Theta}(\mathbf{k})$*, ex post jointly inefficient regulation for* $\hat{\Theta}(\mathbf{k}) < \theta < \Theta^J(\mathbf{k})$*, and ex post efficient regulation for* $\theta \geq \Theta^J(\mathbf{k})$*. Hence, there will be domestic, but no joint regulatory chill.*

In a Pareto optimal agreement, the threshold for regulation satisfies $\hat{\Theta}(\mathbf{k}) \leq \Theta^J(\mathbf{k})$. Hence, there will be regulation whenever it is ex post jointly optimal to regulate. We explained the intuition for this result in the discussion following Proposition 3. Basically, an agreement can implement efficient regulation for $\theta > \Theta^J(\mathbf{k})$ without affecting investment incentives or industry profit by setting compensation payments equal to $\Pi^h(\mathbf{k})$ for all investors and all shocks in the range $\theta > \Theta^J(\mathbf{k})$. The increase in regulatory efficiency benefits the host country.

It is easy to see how domestic regulatory chill can be perceived by the host country as a failure of the agreement. Countries do not know their exact future regulatory needs when they enter into an agreement. It is therefore possible that an agreement that was beneficial in expectation, turns out to be harmful ex post. The harm can materialize as domestic regulatory chill when the host country does not intervene for the fear of compensation payments. Alternatively, and perhaps even more politically provocatively, the host country might choose to regulate and pay compensation. It might then appear as if the agreement forces the host country to pay in order to be able to pursue policies that are desirable from a national perspective. Indeed, in our model, host countries are actually punished for doing what is right: Recall that compensation is paid in equilibrium only for shocks $\theta > \Theta^J(\mathbf{k})$ when regulation is ex post jointly optimal.[38] This criticism of course fails to account for the fact that the increase in investment induced by the agreement would have been valuable under a less severe regulatory shock. Positive investment externalities that occur for shocks $\theta \leq \Theta(\mathbf{k})$, when it would anyway never be in the host country's interest to regulate, can be more than enough to compensate in an ex ante sense for the expected cost of domestic regulatory chill.

[38] This property is similar to how agents are punished in standard moral hazard models, despite exerting the principal's preferred effort.

Lobbying or corruption as a cause of regulatory chill In the discussion of regulatory chill above, the objective function of the host country policy maker is used both to derive the regulatory behavior of the host country, and for evaluating efficiency. This is a natural approach for an economic analysis of basic features of investment agreements. It is also a standard assumption, for instance, in the literature on trade agreements. But there are of course circumstances where the utility $V(\mathbf{k},\theta)$ of the host country government would to be a poor representation of national surplus $W(\mathbf{k},\theta)$ associated with foreign investment, such as when the government responds to lobbying by special interests, or is outright corrupt. To capture implications of such phenomena, assume in standard fashion that the host country government maximizes $V(\mathbf{k},\theta) \equiv W(\mathbf{k},\theta) + \gamma\Pi(\mathbf{k})$, where $\gamma \in [0,1)$ represents the weight attached to foreign operating profits in the decision to regulate. The latter can reflect, for instance, lobbying efforts by foreign investors, or the share of profits that foreign investors pay in terms of bribes to government officials for protection against regulation.[39] The net industry operating profits are thus $(1-\gamma)\Pi(\mathbf{k})$. Importantly, payment for protection is presumed not to be a source of national surplus.

This modification of our setting will affect the implications of host country behavior. $\Theta^J(\mathbf{k})$ still characterizes the ex post efficient level of regulation from the viewpoint of the *negotiating governments*, whereas the efficient threshold for regulation $\Theta^W(\mathbf{k},\gamma) \le \Theta^J(\mathbf{k})$ from a *total surplus* point of view is defined by

$$W(\mathbf{k},\Theta^W) + \Pi(\mathbf{k}) \equiv 0 \Leftrightarrow V(\mathbf{k},\Theta^W) + (1-\gamma)\Pi(\mathbf{k}) \equiv 0$$

for $W(\mathbf{k},\bar{\theta}) + \Pi(\mathbf{k}) \le 0$, and by $\Theta^W(\mathbf{k},\gamma) = \bar{\theta}$ otherwise.

The essential difference between this setting and the one above ($\gamma = 0$) is that the possibility of extracting rents can induce the host country government to *underregulate* from a total surplus viewpoint. As before, let $\hat{\mathbf{k}}$ be the equilibrium investment portfolio under the carve-out mechanism $\mathbf{T}^C$, denote by $\hat{\theta}^C \equiv \Theta^C(\hat{\mathbf{k}})$ the corresponding equilibrium investment protection, let $\hat{\theta}^J \equiv \Theta^J(\hat{\mathbf{k}})$ be the jointly efficient threshold for regulation from the viewpoint of the governments, and $\hat{\theta}^W \equiv \Theta^W(\hat{\mathbf{k}},\gamma)$ the efficient threshold for regulation from a total surplus perspective. There will thus be total surplus-reducing regulatory chill under this agreement if the host country allows production for some shocks $\theta > \hat{\theta}^W$. The following characterization follows directly from the properties of carve-out compensation:

Proposition 9 *If an agreement is based on a carve-out scheme $\mathbf{T}^C$, and the host country government appropriates a share $\gamma > 0$ of foreign operating profits in return for allowing production:*
(i) There will be total surplus-reducing regulatory chill from any agreement with extensive investment protection ($\hat{\theta}^C > \hat{\theta}^J$).
(ii) Lobbying does not cause total surplus-reducing regulatory chill if the agreement has sufficiently

[39] Similar formalizations of lobbying have been used extensively, for instance in the trade policy literature. A recent application is by Maggi and Ossa (2019) in their study of deep trade agreements.

large carve-outs from compensation requirements ($\hat{\theta}^C \leq \hat{\theta}^W$).

By the nature of carve-out compensation, there is only compensation if $\hat{\theta}^C > \hat{\theta}^J$. This means that any agreement with equilibrium compensation payments necessarily suffers from total surplus-reducing regulatory chill if the host country has been influenced by direct lobbying efforts of foreign investors in the negotiation of the agreement ($\gamma > 0$).[40] Hence, compensation payments signal total surplus-reducing regulatory chill in this setting. Finally, note that there can also be total surplus-reducing regulatory chill even absent compensation payments. This occurs if $\hat{\theta}^W < \hat{\theta}^C < \hat{\theta}^J$ and $\theta \in (\hat{\theta}^W, \hat{\theta}^C]$.

Janeba (2019) formalizes the concept of regulatory chill and analyzes its properties in a model with imperfect enforcement of an exogenously specified investment agreement. His main result is that there will be domestic, but no joint regulatory chill. Compared to Janeba (2019), we show that the absence of joint regulatory chill is a general property of Pareto optimal investment agreements if the host country maximizes national surplus. This conclusion does not depend on perfect enforcement of the agreement. To see this, observe that transfers can be set in such a way that each firm receives its *expected* (ex ante of dispute settlement) operating profit in compensation for regulation for shocks $\theta > \Theta^J(\mathbf{k})$, and if enforcement is incomplete. All firms then are indifferent between being regulated or allowed to produce. Regulation will occur because the expected compensation payments are smaller than the host country utility of allowing production. Hence, there will be regulation for all $\theta > \Theta^J(\mathbf{k})$. Imperfect enforcement would be interesting to study in depth in a setting with lobbying ($\gamma > 0$). In particular, imperfect enforcement can reduce the problem of total surplus-reducing regulatory chill because there will be regulation for some $\theta \leq \Theta^J(\mathbf{k})$ if the agreement builds on carve-out compensation $\mathbf{T}^C$. However, we leave this issue for future research.

6.2 North-South versus North-North agreements

The previous analysis has considered a setting where host countries lack the ability to make credible unilateral commitments to compensate foreign investors in case of regulation, and in most instances where investments flow in one direction only. This seems descriptive of the setting for the negotiations regarding a traditional bilateral investment agreement between a developed and a developing country. Such agreements were (and to some extent still are) formed with the primary purpose of overcoming weaknesses in the legal institutions of the developing country, in order to stimulate investment flows from the developed to the developing country. We refer to these as *North-South agreements*. But this setting does not appropriately describe the context of agreements between developed economies, such as the agreement that would have resulted from the TTIP negotiations between the EU and the US, or the recent EU-Canada agreement. These economies are largely capable of making *credible unilateral commitments* to protect incoming foreign investment through

[40] Observe that foreign firms always have an *indirect* influence over the construction of the agreement by our assumption that the source country has bargaining power and maximizes the expected profit of foreign investment.

their domestic legal and regulatory frameworks. Additionally, the agreements are meant to stimulate investment flows in *both* directions. We will refer to these agreements *North-North agreements.* These differences between the two types of settings will have important implications for the scope of the agreements and the distribution of the resulting surplus, as we shall see.

6.2.1 The different rationales of the agreements

To see the importance of South's lack of commitment for ability for the scope of a North-South agreement, suppose South can commit to investment protection even without an agreement, but investment can only flow from North to South. Also, to facilitate the comparison, let us assume that unilateral investment commitments have the same qualitative feature as investment agreements in that they build on carve-out compensation. Absent an agreement, South then chooses the carve-out scheme $\mathbf{T}^U$ that maximizes $\tilde{V}(\mathbf{T})$, where $\tilde{V}(\mathbf{T})$ now is interpreted as the expected national surplus. Let the equilibrium expected surpluses for the two countries be denoted $v^U \equiv \tilde{V}(\mathbf{T}^U)$ and $\pi^U \equiv \tilde{\Pi}(\mathbf{T}^U)$, and let $\omega^U \equiv v^U + \pi^U$ be the expected total surplus. Finally, denote by θ^U the equilibrium investment protection under $\mathbf{T}^U$, and assume that the expected industry profit is strictly increasing in investment protection.

South cannot possibly benefit from entering into an agreement with North, even if it could dictate the terms of the agreement, since it can *unilaterally ensure the maximal surplus v^U that it can obtain from an agreement.* This property holds even if there are aggregate gains from trade, i.e., there is an agreement $\mathbf{T}^C$ such that $\tilde{V}(\mathbf{T}^C) + \tilde{\Pi}(\mathbf{T}^C) > \omega^U$. All such negotiations would necessarily fail because there are no unilateral gains for the host country under one-way investment (absent side payments). It follows that the role of North-South investment agreements is to *gain access to the credible enforcement mechanisms* that support the agreements. This role corresponds closely to the notion of trade agreements as *commitment devices* that help governments withstand domestic protectionist pressures. But note also that investment agreements, while inducing investment from North, are still imperfect substitutes for credible domestic legislation from South's perspective, because South will generally have to share the net gains from an agreement with North.

The purpose of a North-North agreement must clearly differ, since these countries by assumption are able to make credible unilateral commitments. But there is still a role for investment agreements, stemming from the *level* at which unilateral commitments are made. If we let Home and Foreign be the two countries in the North-North relationship, then the total expected surplus generated in Home is $\tilde{V}(\mathbf{T}^U) + \tilde{\Pi}(\mathbf{T}^U)$ absent an investment agreement. By ignoring the positive effect of investment protection on the industry profit $\tilde{\Pi}(\mathbf{T}^U)$ of Foreign investment, θ^U is too small from a total surplus perspective. The same is true for the equilibrium investment protection θ^{*U} in Foreign under the carve-out compensation scheme $\mathbf{T}^{*U}$ that maximizes expected national surplus $\tilde{V}^*(\mathbf{T}^*)$ in Foreign. Absent any agreement, there will be too weak protection of foreign investment in both countries by their *disregard of the benefits of their respective investment protection commitments* *for foreign investors.* Consequently, there is scope for an agreement that *coordinates an exchange of investment protection commitments* and induces countries to internalize positive international externalities from their domestic protection regimes. This argument parallels the standard view of the role of trade agreements, which sees these agreements as solutions to Prisoners' Dilemmas that allow countries to exchange mutually beneficial tariff concessions.[41] In sum:

Observation 3 *Concerning the scope for an investment agreement:*
(1) A host country that can unilaterally implement any compensation mechanism to protect investment through its domestic legal system will not enter into an agreement over one-way investment, even if this would increase the total surplus.
(2) With two-way investment flows there is scope for an agreement regardless of unilateral commitment abilities.
(3) The rationale for a traditional bilateral North-South agreement is the latter country's lack of unilateral commitment possibilities regarding investment protection.
(4) The rationale for a North-North agreement is to coordinate investment protection so as to internalize positive external effects from national investment protection.

6.2.2 The different distributional impacts of the agreements

In a North-South agreement, both parties agree to increase investment protection in South at least up to θ^U. As long as neither party can dictate the terms of the agreement, the negotiated investment protection will be strictly above θ^U, but not so high that South would not benefit from the agreement. Hence, the negotiated agreement satisfies $\tilde{V}(\mathbf{T}^C) > v^0$. For the parties to a North-North agreement, there are two sources of surplus. In Home, the first source is the expected national surplus $\tilde{V}(\mathbf{T}^C)$ associated with inward investment from Foreign. The second source is the expected industry profit $\tilde{\Pi}^*(\mathbf{T}^{*C})$ of outward investment into Foreign. Home and Foreign have a joint incentive to let investment protection be at least at the unilaterally optimal levels (θ^U, θ^{*U}), but investment protection will be strictly higher if both countries have bargaining power. Home's net gain from the negotiated agreement will then be $\tilde{V}(\mathbf{T}^C) + \tilde{\Pi}^*(\mathbf{T}^{*C}) - v^U - \pi^{*U} > 0$, where $\pi^{*U} = \tilde{\Pi}^*(\mathbf{T}^{*U})$ is the expected industry profit from Home's outward investment into Foreign if there is no agreement, and Foreign commits to carve-out compensation $\mathbf{T}^{*U}$. By implication:

$$\tilde{\Pi}^*(\mathbf{T}^{*C}) - \pi^{*U} > v^U - \tilde{V}(\mathbf{T}^C) > 0. \tag{13}$$

Although both parties yield concessions with regard to protection of inward investment under a North-North agreement, these losses are outweighed by the increases in expected profits from out-

[41] A significant fraction of bilateral investment agreements are nowadays between developing countries. Such *South-South agreements* are often formed between countries that lack capacity to make credible unilateral undertakings. But the more symmetric nature of the contracting parties, suggests that the agreements are meant to promote two-way investment. In these cases, there is a double benefit to the agreements: improved enforcement through arbitration courts, and internalization of international policy externalities.

ward investment. We collect these results in the following proposition:

Proposition 10 *Concerning the distribution of surplus in investment agreements:*
(1) North-South and North-North agreements entail more investment protection than the levels (θ^U, θ^{*U}) *that maximize host country expected national surplus.*
(2) A North-South agreement benefits investors from North and increases expected national surplus in South.
(3) A North-North agreement benefits foreign investors in both countries, but reduces expected national surplus in both countries.

We believe that these results shed light on the policy debate regarding investment agreements. The costs and benefits for the Southern parties to North-South agreements have been discussed for years. But several thousands such agreements were signed without much political opposition. This contrasts sharply with the heated debate concerning the attempts to include investment protection in North-North agreements, and most notably in CETA, TPP, and TTIP. Existing legal systems in the EU and the US are in all likelihood capable of providing sufficient protection of foreign investment to internalize national surplus effects. Our framework suggests that the additional investment protection offered by these agreements would mainly benefit foreign investors and harm the rest of society. These agreements will thus *always appear too protective of foreign investor interests, from the point of view of the rest of society.*[42]

Konrad (2017) considers the distribution effects of an agreement in a model in which firms have a strategic incentive to invest in order to reduce the likelihood of environmental regulation. Increased investment protection from an exogenously specified agreement benefits investors, but exacerbates an already existing overinvestment and underregulation problem. Hence, the host country is worse off compared to a situation when there was no agreement. Total surplus can go down as a result of the agreement. In Schjelderup and Stähler (2019), firms overinvest because of market power, whereas environmental regulation is exogenous. Firms gain from investment protection, and the authors establish sufficient conditions for when the effects on the host country are negative. In our model, an agreement with one-way investment flows cannot cause national surplus to decrease if the objective of the host country is to maximize national surplus. We consider agreements with endogenous contract terms that must fulfill both countries' individual participation constraints. However, negotiated agreements can reduce national surplus under two-way investment flows, as we have seen.

[42] We note that the trade-off that we here identify is also recognized in a report to the Parliament of Australia (2016, p. 64) on TPP, which states that "[u]nder the TPP ISDS provisions, Australian investors have more to gain than the Australian Government and the Australian people have to lose." This is precisely the message of equation (13).

7 Concluding remarks

International investment agreements are economy-wide treaties sustained by highly potent enforcement mechanisms that protect foreign investors against a wide array of host country policy interventions. Severe criticism has been directed against these agreements by academics, politicians and the general public. Yet, the previous economic literature has offered very little guidance how to interpret these agreements and how to understand the controversies surrounding them. There hardly exists any economic theory regarding fundamental aspects of the agreements, such as their rationale, efficiency properties, and distributional effects. The purpose of this paper has been to contribute to filling this void. To this end, we have examined negotiated investment agreements that share core features with actual agreements. This approach has generated a number of new results that can help explain how investment agreements function, and that also shed light on the validity of main arguments in the policy debate.

For instance, we have shown that a negotiated agreement based on carve-out compensation implements the jointly efficient outcome in a robust set of circumstances, when investment and regulation incentives are distorted, and there are conflicts of interests with regard to the distribution of the surplus from the agreement. The analysis thus provides an economic foundation for the legal principle of full compensation, and for allowing uncompensated interventions to achieve certain policy objectives. We have identified fundamental differences between agreements with one- and two-way investment flows, and we have demonstrated the importance of unilateral commitment capacity for the role and distributional impact of investment agreements. We have provided a general argument for why Pareto optimal agreements yield domestic, but not joint, regulatory chill if the host country maximizes national surplus. We have examined how compensation requirements for regulatory expropriation interact with other core provisions in investment agreements. Negotiated compensation schemes either entail complete investment protection against direct expropriation or offer the same investment protection for all types of expropriation; agreements only featuring National Treatment are Pareto dominated by those with carve-out compensation; and Pareto optimal agreements include ISDS. These properties are consistent with standard features of actual investment agreements.

The literature on investment agreements is just beginning to emerge. We conclude by pointing to some aspects of these agreements that still have no explanation.

First, 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 they reflect an exchange of concessions in the investment and trade areas (Maggi, 2016). But the precise form of interaction between investment and trade undertakings remains to be identified.

Second, we have considered formation of an investment agreement between a pair of countries, without taking into account interactions with other countries. Several important aspects of the investment regime have thus been neglected. For instance, a striking feature of international investment protection is the lack of a multilateral investment agreement similar to the World Trade Organization. It is also relevant to think of parallel negotiations where a developed country simultaneously negotiates investment agreements with developing countries. Is there a race to the bottom concerning investment protection? To account for such interactions, one would need to consider interrelated negotiations and the sequential formation of agreements.

Third, we have assumed that the only effect of an agreement for the source country under one-way investment flows, is to increase the expected profits of outward investments. However, such an agreement can redirect investments from the source to the host country by reducing the barriers to foreign investments. It would be interesting to examine the consequences of such source country effects on the negotiated agreement.

Fourth, carve-outs have both a quantitative and qualitative dimension. Our framework captures the quantitative dimension by specifying carve-outs for all regulatory shocks above a threshold. The qualitative dimension of carve-outs is that they often apply to specific policy objectives such as the protection of human, animal and plant life and health. It is also increasingly common that certain industries do not receive any investment protection from the agreement. To capture such qualitative aspects, the analysis should include multidimensional regulatory shocks.

Fifth, we have left out arbitration from most of the analysis by assuming that agreements are perfectly enforceable, although we did verify that key results do not depend crucially on perfect enforcement. The economic literature on dispute settlement in investment agreements has analyzed implications of arbitration courts receiving noisy signals about the true state of the world. Many issues concerning dispute settlement in investment agreements have yet to be investigated. For instance, a core issue is how to interpret the notion of investors' *legitimate expectations* regarding regulation. This concept plays a central role in many agreements and in case law to determine whether regulation should be compensated, but does not seem to have any obvious economic meaning. It would be valuable to endogenize dispute settlement and analyze arbitration in greater detail in the analysis of investment agreements.

References

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

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

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

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

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.

Falvey, Rod and Neil Foster-McGregor (2015). "North-South FDI and Bilateral Investment Treaties." UNUMERIT Working Papers No. 010.

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

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

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

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

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

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

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

Kohler, Wilhelm and Frank Stähler (2019). "The Economics of Investor Protection: ISDS versus National Treatment." Mimeo, April 1.

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

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

Maggi, Giovanni and Ralph Ossa (2019). "Are Trade Agreements Good for You?" Draft.

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.

Ossa, Ralph, Robert W. Staiger and Alan O. Sykes (2019): "Disputes in International Investment and Trade." Preliminary draft, March.

Parliament of Australia (2016): Trans-Pacific Partnership Agreement, Joint Standing Committee on Treaties. Report 165.

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

Schjelderup, Guttorm and Frank Stähler (2019). "Investor State Dispute Settlement and Multinational Firm Behavior." Mimeo, July 10.

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 (2018). "An Optimal Investor State Dispute Settlement Mechanism." Mimeo, March 1.

Sykes, Alan O. (2005). "Public vs. Private Enforcement of International Economic Law: Of Standing and Remedy." John M. Olin Law & Economics Working Paper, No. 235, University of Chicago Law School.

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

Vandevelde, Kenneth J. (1998). "The Political Economy of a Bilateral Investment Treaty." *American Journal of International Law* 92(4), 621-641.

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

A Appendix

We start by deriving some intermediary results concerning regulation under investment agreements.

Lemma 1 *Consider an investment agreement based on general compensation* $\mathbf{T}$*. Let* $M(\mathbf{k})$ *[*$M^r(\mathbf{k})$*] be the subset of shock realizations for which it is [strictly] optimal for the host country to allow production [regulate] for arbitrary investment portfolio* $\mathbf{k}$*:*

$$\begin{aligned} M(\mathbf{k}) \equiv & \quad \{\theta \in [\underline{\theta}, \bar{\theta}] : V(\mathbf{k},\theta) + \textstyle\sum_{h=1}^{H} T^h(\mathbf{k},\theta) \geq 0\}, \\ M^r(\mathbf{k}) \equiv & \quad \{\theta \in [\underline{\theta}, \bar{\theta}] : \theta \notin M(\mathbf{k})\}. \end{aligned} \tag{A.1}$$

The investment agreement (weakly) reduces regulation compared to the case of no agreement: $[\underline{\theta}, \Theta(\mathbf{k})] \subset M(\mathbf{k})$*.*

Proof: The assumptions that $V(\mathbf{k},\theta)$ is strictly decreasing in θ and $T^h(\mathbf{k},\theta) \geq 0$ for all $(\mathbf{k},\theta)$ jointly imply

$$V(\mathbf{k},\theta) + \textstyle\sum_{h=1}^{H} T^h(\mathbf{k},\theta) \geq V(\mathbf{k},\theta) \geq V(\mathbf{k},\Theta(\mathbf{k})) = 0$$

for all $\theta \leq \Theta(\mathbf{k})$, and therefore $[\underline{\theta}, \Theta(\mathbf{k})] \subset M(\mathbf{k})$.■

The characterizations of $M(\mathbf{k})$ and $M^r(\mathbf{k})$ are particularly simple under carve-out compensation because the host country internalizes the full effects of its decisions for all shocks $\theta \leq \Theta^C(\mathbf{k})$:

Lemma 2 *Consider an investment agreement based on carve-out compensation* $\mathbf{T}^C$ *and investment protection* $\Theta^C(\mathbf{k}) \geq \Theta(\mathbf{k})$.[43] *In this case,* $M(\mathbf{k}) = [\underline{\theta}, \hat{\Theta}(\mathbf{k})]$*, where* $\hat{\Theta}(\mathbf{k}) = \min\{\Theta^C(\mathbf{k}); \Theta^J(\mathbf{k})\}$ *characterizes the threshold for regulation under* $\mathbf{T}^C$*. Firm h's expected profit equals:*

$$F(\Theta^C(\mathbf{k}))\Pi^h(\mathbf{k}) - R^h(k_k). \tag{A.2}$$

Proof: For $\theta \leq \hat{\Theta}(\mathbf{k})$, the net benefit of allowing production is non-negative:

$$V(\mathbf{k},\theta) + \textstyle\sum_{h=1}^{H} T^{hC}(\mathbf{k},\theta) = V(\mathbf{k},\theta) + \Pi(\mathbf{k}) \geq V(\mathbf{k},\Theta^J(\mathbf{k})) + \Pi(\mathbf{k}) \geq 0.$$

If $\hat{\Theta}(\mathbf{k}) = \Theta^C(\mathbf{k})$ and $\theta > \hat{\Theta}(\mathbf{k})$, then it is strictly optimal to regulate:

$$V(\mathbf{k},\theta) + \textstyle\sum_{h=1}^{H} T^{hC}(\mathbf{k},\theta) = V(\mathbf{k},\theta) < V(\mathbf{k},\hat{\Theta}(\mathbf{k})) \leq V(\mathbf{k},\Theta(\mathbf{k})) \leq 0.$$

If $\hat{\Theta}(\mathbf{k}) = \Theta^J(\mathbf{k})$ and $\theta > \hat{\Theta}(\mathbf{k})$, then it is also strictly optimal to regulate:

$$V(\mathbf{k},\theta) + \textstyle\sum_{h=1}^{H} T^{hC}(\mathbf{k},\theta) \leq V(\mathbf{k},\theta) + \Pi(\mathbf{k}) < V(\mathbf{k},\Theta^J(\mathbf{k})) + \Pi(\mathbf{k}) = 0.$$

[43] The case with $\Theta^C(\mathbf{k}) < \Theta(\mathbf{k})$ in $\mathbf{T}^C$ is uninteresting because the threshold for regulation in that case is $\Theta(\mathbf{k})$, and $\Theta^C(\mathbf{k})$ therefore is non-binding.

The firm obtains its operating profit for all $\theta \leq \Theta^C(\mathbf{k})$ regardless of whether it is allowed to produce or not. It is regulated for all $\theta > \Theta^C(\mathbf{k})$ by $\Theta^C(\mathbf{k}) \geq \hat{\Theta}(\mathbf{k})$. Regulation is uncompensated in this case by the properties of $\mathbf{T}^C$.■

A.1 Proof of Proposition 1

Consider first the possibility of implementing $(\mathbf{k}^J, \theta^J)$ through a carve-out compensation mechanism:

Lemma 3 *An investment agreement with carve-out compensation* $\mathbf{T}^C$ *can implement the jointly efficient outcome* $(\mathbf{k}^J, \theta^J)$ *if and only if*

$$\Pi^h(\mathbf{k}^J) - R^h(k_h^J) \geq \pi_h^d(\mathbf{k}_{-h}^J) \equiv \max_{k_h \geq 0}[F(\Theta(k_h, \mathbf{k}_{-h}^J))\Pi^h(k_h, \mathbf{k}_{-h}^J) - R^h(k_h)] \text{ for all } h \in \mathcal{H}. \quad \text{(A.3)}$$

Proof: We first consider sufficiency. Condition (A.3) is equivalent to

$$\Xi(\mathbf{k}^J) \equiv \max_{h \in \mathcal{H}}\{\frac{\pi_h^d(\mathbf{k}_{-h}^J) + R^h(k_h^J)}{\Pi^h(\mathbf{k}^J)}\} \leq 1.$$

Consider a $\mathbf{T}^C$ where investment protection $\theta^{JC} \equiv \Theta^C(\mathbf{k}^J)$ under efficient investment satisfies

$$\max\{\Xi(\mathbf{k}^J); F(\theta^J)\} \leq F(\theta^{JC}) \leq 1, \quad \text{(A.4)}$$

and where $\Theta^C(\mathbf{k}) \equiv \Theta(\mathbf{k})$ for all $\mathbf{k} \neq \mathbf{k}^J$. The expected profit for h of choosing k_h^J equals $F(\theta^{JC})\Pi^h(\mathbf{k}^J) - R^h(k_h^J)$ if all other firms choose $\mathbf{k}_{-h}^J$, by $\theta^{JC} \geq \Theta(\mathbf{k}^J)$ and Lemma 2. If h unilaterally deviates to $k_h \neq k_h^J$, then its expected profit becomes instead

$$\begin{aligned} F(\Theta(k_h, \mathbf{k}_{-h}^J))\Pi^h(k_h, \mathbf{k}_{-h}^J) - R^h(k_h) &\leq \pi_h^d(\mathbf{k}_{-h}^J) = \left(\frac{\pi_h^d(\mathbf{k}_{-h}^J) + R^h(k_h^J)}{\Pi^h(\mathbf{k}^J)}\right)\Pi^h(\mathbf{k}^J) - R^h(k_h^J) \\ &\leq \Xi(\mathbf{k}^J)\Pi^h(\mathbf{k}^J) - R^h(k_h^J) \leq F(\theta^{JC})\Pi^h(\mathbf{k}^J) - R^h(k_h^J). \end{aligned}$$

As unilateral deviations are unprofitable for all firms, it follows that $\mathbf{T}^C$ with the stated investment protection implements $\mathbf{k}^J$. This $\mathbf{T}^C$ mechanism implements also efficient regulation for the equilibrium investment portfolio $\mathbf{k}^J$ by $\theta^{JC} \geq \theta^J$ and Lemma 2.

We next prove necessity by showing that (A.3) is necessary to implement $(\mathbf{k}^J, \theta^J)$ for *any* $\mathbf{T}$ required to satisfy $0 \leq T^h(\mathbf{k}, \theta) \leq \Pi^h(\mathbf{k})$ for all h and $(\mathbf{k}, \theta)$. For such $\mathbf{T}$,

$$\Pi^h(\mathbf{k}^J)\textstyle\int_{M(\mathbf{k}^J)}dF(\theta) + \int_{M^r(\mathbf{k}^J)}T^h(\mathbf{k}^J, \theta)dF(\theta) - R^h(k_h^J) \leq \Pi^h(\mathbf{k}^J) - R^h(k_h^J)$$

is the expected profit of firm h if all firms invest efficiently, and where the inequality follows from $T^h(\mathbf{k}^J, \theta) \leq \Pi^h(\mathbf{k}^J)$. In other words, $\Pi^h(\mathbf{k}^J) - R^h(k_h^J)$ represents an upper bound to what firm h can earn by investing the intended k_h^J. The expected profit for h of instead investing the k_h^d that enters into $\pi^d_h(\mathbf{k}^J_{-h})$ is

$$
\begin{aligned}
& \Pi^h(k^d_h,\mathbf{k}^J_{-h})\textstyle\int_{M(k^d_h,\mathbf{k}^J_{-h})}dF(\theta)+\int_{M^r(k^d_h,\mathbf{k}^J_{-h})}T^h(k^d_h,\mathbf{k}^J_{-h},\theta)dF(\theta)-R^h(k^d_h) \\
= \;& \Pi^h(k^d_h,\mathbf{k}^J_{-h})[\textstyle\int_{M(k^d_h,\mathbf{k}^J_{-h})}dF(\theta)-F(\Theta(k^d_h,\mathbf{k}^J_{-h}))]+\int_{M^r(k^d_h,\mathbf{k}^J_{-h})}T^h(k^d_h,\mathbf{k}^J_{-h},\theta)dF(\theta)+\pi^d_h(\mathbf{k}^J_{-h}) \\
\geq \;& \pi^d_h(\mathbf{k}^J_{-h}).
\end{aligned}
$$

The first term on the second row is non-negative by $\int_{M(\mathbf{k})}dF(\theta)\geq F(\Theta(\mathbf{k}))$; see Lemma 1. The second term is non-negative by $T^h(\mathbf{k},\theta)\geq 0$. In other words, firm h can earn at least $\pi^d_h(\mathbf{k}^J_{-h})$ under a deviation from k^J_h under any compensation mechanism with non-negative compensation. A deviation is strictly profitable for at least one firm if (A.3) is violated.■

Having established the necessary and sufficient condition (A.3) for $\mathbf{T}^C$ to be able to implement $(\mathbf{k}^J,\theta^J)$, let us now turn to negotiations. The individual participation constraints

$$\textstyle\int_{\underline{\theta}}^{\theta^J}V(\mathbf{k}^J,\theta)dF(\theta)\geq v^0,\ \Pi(\mathbf{k}^J)-R(\mathbf{k}^J)\geq\pi^0 \tag{A.5}$$

immediately follow. If the first constraint is violated, the host country strictly prefers to be without an agreement rather than accept *any* general compensation scheme that implements $(\mathbf{k}^J,\theta^J)$ because expected compensation payments are non-negative: $\tilde{V}(\mathbf{T})<v^0$. If the second constraint is violated, the source country strictly prefers to be without an agreement rather than accept *any* compensation scheme that implements $(\mathbf{k}^J,\theta^J)$ and entails non-punitive damages $T^{Nh}(\mathbf{k},\theta)\in[0,\Pi^h(\mathbf{k})]$ for all h: $\tilde{\Pi}(\mathbf{T}^N)<\pi^0$. Assume that (A.3) and (A.5) hold, and that $F(\theta^{JC})$ characterized in (10) is contained in $[\max\{\Xi(\mathbf{k}^J));F(\theta^J)\},1]$. Substituting $F(\theta^{JC})$ into the Nash Product defined in (6) yields $\mathcal{N}(\mathbf{T}^C)=\alpha^\alpha(1-\alpha)^{1-\alpha}(\omega^J-\omega^0)$ after simplification, where $\omega^J\equiv\Omega(\mathbf{k}^J)=\max_{\mathbf{k}}\Omega(\mathbf{k})$, and

$$\Omega(\mathbf{k})\equiv\textstyle\int_{\underline{\theta}}^{\Theta^J(\mathbf{k})}[V(\mathbf{k},\theta)+\Pi(\mathbf{k})]dF(\theta)-R(\mathbf{k}) \tag{A.6}$$

is the expected joint surplus under ex post jointly optimal regulation for arbitrary $\mathbf{k}$. Now define the unconstrained Nash Product

$$\hat{\mathcal{N}}(\mathbf{T},s)\equiv[\tilde{V}(\mathbf{T})-s-v^0]^\alpha[\tilde{\Pi}(\mathbf{T})+s-\pi^0]^{1-\alpha}$$

for an agreement with some general compensation scheme $\mathbf{T}$, where s is an unconstrained side-payment. $\hat{\mathcal{N}}(\mathbf{T},s)$ is strictly quasi-concave in s and reaches its global optimum at

$$S(\mathbf{T})=(1-\alpha)(\omega(\mathbf{T})-\omega^0)-\tilde{\Pi}(\mathbf{T})+\pi^0,$$

where

$$\omega(\mathbf{T})\equiv\tilde{V}(\mathbf{T})+\tilde{\Pi}(\mathbf{T})=\textstyle\int_{M(\hat{\mathbf{k}})}[V(\hat{\mathbf{k}},\theta)+\Pi(\hat{\mathbf{k}})]dF(\theta)-R(\hat{\mathbf{k}}) \tag{A.7}$$

is the expected joint surplus under $\mathbf{T}$, and $\hat{\mathbf{k}}$ is the equilibrium investment. By optimality of $S(\mathbf{T})$:

$$\hat{\mathcal{N}}(\mathbf{T}, S(\mathbf{T})) = \alpha^{\alpha}(1-\alpha)^{1-\alpha}(\omega(\mathbf{T}) - \omega^0) \geq \hat{\mathcal{N}}(\mathbf{T}, 0) = \mathcal{N}(\mathbf{T}).$$

We finally demonstrate $\omega^J \geq \omega(\mathbf{T})$, so that we can combine inequalities to get $\mathcal{N}(\mathbf{T}^C) \geq \mathcal{N}(\mathbf{T})$. $\omega^J \geq \omega(\mathbf{T})$ follows from $\omega^J \geq \Omega(\hat{\mathbf{k}})$ and

$$\Omega(\hat{\mathbf{k}}) - \omega(\mathbf{T}) = \int_{M^r(\hat{\mathbf{k}}) \cap [\underline{\theta}, \Theta^J(\hat{\mathbf{k}})]} [V(\hat{\mathbf{k}},\theta) + \Pi(\hat{\mathbf{k}})] dF(\theta) - \int_{M(\hat{\mathbf{k}}) \cap (\Theta^J(\hat{\mathbf{k}}), \bar{\theta}]} [V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}})] dF(\theta) \geq 0$$

because regulation is ex post optimal if $\Theta^J(\hat{\mathbf{k}})$ defines the threshold for regulation. We conclude:

Proposition A.1 *An investment agreement with carve-out compensation* $\mathbf{T}^C$ *implements the jointly efficient outcome* $(\mathbf{k}^J,\theta^J)$ *and maximizes the Nash Product* $\mathcal{N}(\mathbf{T})$ *in the family of general compensation schemes* $\mathbf{T}$ *if the participation constraints (A.5) and following conditions are met:* $\Xi(\mathbf{k}^J) \leq 1$ *and*

$$\max\{\Xi(\mathbf{k}^J); F(\theta^J)\} \leq \frac{R(\mathbf{k}^J) + \pi^0}{\Pi(\mathbf{k}^J)} + (1-\alpha)\frac{\omega^J - \omega^0}{\Pi(\mathbf{k}^J)} \leq 1. \tag{A.8}$$

Robustness The conditions of Proposition A.1 are satisfied under a robust set of circumstances. Assume that $\Xi(\mathbf{k}^J) < 1$ and $F(\theta^J) < 1$ and that (A.5) hold strictly. As $\frac{R(\mathbf{k}^J)+\pi^0}{\Pi(\mathbf{k}^J)} < 1$, a necessary and sufficient condition for (A.8) to hold for a non-degenerate interval $\alpha \in [\underline{\alpha}, \bar{\alpha}]$, $0 \leq \underline{\alpha} < \bar{\alpha} \leq 1$ is

$$\frac{R(\mathbf{k}^J) + \pi^0}{\Pi(\mathbf{k}^J)} + \frac{\omega^J - \omega^0}{\Pi(\mathbf{k}^J)} > \max\{\Xi(\mathbf{k}^J); F(\theta^J)\}.$$

If $\Xi(\mathbf{k}^J) \leq F(\theta^J)$, then

$$\frac{R(\mathbf{k}^J) + \pi^0}{\Pi(\mathbf{k}^J)} + \frac{\omega^J - \omega^0}{\Pi(\mathbf{k}^J)} - \max\{\Xi(\mathbf{k}^J); F(\theta^J)\} = \frac{1}{\Pi(\mathbf{k}^J)}\left(\int_{\underline{\theta}}^{\theta^J} V(\mathbf{k}^J,\theta) dF(\theta) - v^0\right) > 0,$$

and we are done. If $\Xi(\mathbf{k}^J) > F(\theta^J)$, then we can rewrite the above condition as

$$\omega^J - \omega^0 > \Xi(\mathbf{k}^J)\Pi(\mathbf{k}^J) - R(\mathbf{k}^J) - \pi^0. \tag{A.9}$$

To evaluate (A.9), we add more structure. If $H = 1$, then the right-hand side of (A.9) is zero because $\pi_h^d(\mathbf{k}_{-h}^J) = \pi^0$ for $H = 1$, and we are done. If $H \geq 2$ and all firms are symmetric, then (A.9) becomes $\omega^J - \omega^0 > H(\pi^d - \pi^0)$, where π^d here is the deviation profit of a representative firm if all other firms invest k^J, and π^0 is its equilibrium profit absent any agreement. This latter inequality is satisfied if, for instance, $k^J > k^0$ and each firm is better off if the other firms in the industry invest less rather than more when there is no investment agreement, because in this case $\pi^d < \pi^0$. These conditions are sufficient, but not necessary. By continuity, (A.9) holds for $H \geq 2$ also if there is some degree of asymmetry and if $\pi_h^d(\mathbf{k}_{-h}^J) > \pi^0$, but not too large.■

A.2 Proof of Proposition 2

Let $\hat{\mathbf{k}}$ be the equilibrium investment profile under an initial agreement with non-punitive compensation $\mathbf{T}^N$. Define the level of investment protection $\Theta^C(\mathbf{k})$ in an alternative agreement with carve-out compensation $\mathbf{T}^C$ by

$$F(\Theta^C(\mathbf{k})) \equiv \int_{M(\mathbf{k})} dF(\theta) + \int_{M^r(\mathbf{k})} \beta(\mathbf{k},\theta) dF(\theta) \leq 1.$$

Observe that $\Theta^C(\mathbf{k}) \geq \Theta(\mathbf{k})$ because $\int_{M(\mathbf{k})} dF(\theta) \geq F(\Theta(\mathbf{k}))$; see Lemma 1. By Lemma 2, the threshold for regulation under $\mathbf{T}^C$ is $\hat{\Theta}(\mathbf{k}) \equiv \min\{\Theta^C(\mathbf{k}); \Theta^J(\mathbf{k})\}$. All firms therefore have the same expected investment profit under both compensation schemes, and for all $\mathbf{k}$:

$$F(\Theta^C(\mathbf{k}))\Pi^h(\mathbf{k}) - R^h(k_h) = \int_{M(\mathbf{k})} dF(\theta)\Pi^h(\mathbf{k}) + \int_{M^r(\mathbf{k})} \beta(\mathbf{k},\theta)\Pi^h(\mathbf{k}) dF(\theta) - R^h(k_h).$$

Hence, $\hat{\mathbf{k}}$ can be sustained as an equilibrium also under $\mathbf{T}^C$. As expected investment profits are the same for all firms in both agreements, the expected industry profits are identical: $\tilde{\Pi}(\mathbf{T}^C) = \tilde{\Pi}(\mathbf{T}^N)$. Consider next the expected host country utility. As the marginal utility of compensation is constant and the same for all parties, it follows that $\tilde{V}(\mathbf{T}^N) = \omega(\mathbf{T}^N) - \tilde{\Pi}(\mathbf{T}^N)$ and $\tilde{V}(\mathbf{T}^C) = \omega(\mathbf{T}^C) - \tilde{\Pi}(\mathbf{T}^C)$, where $\omega(\mathbf{T})$ denotes the expected total surplus of general compensation $\mathbf{T}$; see (A.7). Hence,

$$\tilde{V}(\mathbf{T}^C) - \tilde{V}(\mathbf{T}^N) = \omega(\mathbf{T}^C) - \omega(\mathbf{T}^N) = \int_{\underline{\theta}}^{\hat{\theta}} [V(\hat{\mathbf{k}},\theta) + \Pi(\hat{\mathbf{k}})] dF(\theta) - \int_{M(\hat{\mathbf{k}})} [V(\hat{\mathbf{k}},\theta) + \Pi(\hat{\mathbf{k}})] dF(\theta),$$

where $\hat{\theta} \equiv \hat{\Theta}(\hat{\mathbf{k}}) = \min\{\hat{\theta}^C; \hat{\theta}^J\}$, $\hat{\theta}^C \equiv \Theta^C(\hat{\mathbf{k}})$, and $\hat{\theta}^J \equiv \Theta^J(\hat{\mathbf{k}})$. Add and subtract $V(\hat{\mathbf{k}},\hat{\theta})$ under each of the two integrals and rewrite:

$$\begin{aligned} \omega(\mathbf{T}^C) - \omega(\mathbf{T}^N) &= \int_{M^r(\hat{\mathbf{k}}) \cap [\underline{\theta},\hat{\theta}]} [V(\hat{\mathbf{k}},\theta) - V(\hat{\mathbf{k}},\hat{\theta})] dF(\theta) + \int_{M(\hat{\mathbf{k}}) \cap (\hat{\theta},\bar{\theta}]} [V(\hat{\mathbf{k}},\hat{\theta}) - V(\hat{\mathbf{k}},\theta)] dF(\theta) \\ &\quad + [V(\hat{\mathbf{k}},\hat{\theta}) + \Pi(\hat{\mathbf{k}})][F(\hat{\theta}) - F(\hat{\theta}^C) + \int_{M^r(\hat{\mathbf{k}})} \beta(\hat{\mathbf{k}},\theta) dF(\theta)]. \end{aligned}$$

The two terms on the first row are non-negative because $V(\hat{\mathbf{k}},\theta)$ is decreasing in θ. The term on the second row is zero if $\hat{\theta}^C > \hat{\theta}^J$ because then $V(\hat{\mathbf{k}},\hat{\theta}) + \Pi(\hat{\mathbf{k}}) = V(\hat{\mathbf{k}},\hat{\theta}^J) + \Pi(\hat{\mathbf{k}}) = 0$. It is non-negative if $\hat{\theta}^C \leq \hat{\theta}^J$ because then $V(\hat{\mathbf{k}},\hat{\theta}) + \Pi(\hat{\mathbf{k}}) \geq V(\hat{\mathbf{k}},\hat{\theta}^J) + \Pi(\hat{\mathbf{k}}) \geq 0$ and $F(\hat{\theta}) - F(\hat{\theta}^C) + \int_{M^r(\hat{\mathbf{k}})} \beta(\hat{\mathbf{k}},\theta) dF(\theta) = \int_{M^r(\hat{\mathbf{k}})} \beta(\hat{\mathbf{k}},\theta) dF(\theta) \geq 0$. Hence, $\omega(\mathbf{T}^C) \geq \omega(\mathbf{T}^N)$, which concludes the proof.■

A.3 Proof of Proposition 3

We prove the result for the case of positive equilibrium investment, $\hat{k}_h > 0$ for some $h \in \mathcal{H}$, because an initial investment agreement $\mathbf{T}$ without any investment is an economically uninteresting benchmark. We first use the threshold function $\hat{\Theta}(\mathbf{k})$ (defined below) to create four partitions of

$[\underline{\theta},\bar{\theta}]$:

$$\begin{aligned} A(\mathbf{k}) &\equiv \{\theta \in M(\mathbf{k}) \cap [\underline{\theta}, \hat{\Theta}(\mathbf{k})]\}, \\ A^r(\mathbf{k}) &\equiv \{\theta \in M^r(\mathbf{k}) \cap [\underline{\theta}, \hat{\Theta}(\mathbf{k})]\}, \\ B(\mathbf{k}) &\equiv \{\theta \in M(\mathbf{k}) \cap (\hat{\Theta}(\mathbf{k}), \bar{\theta}]\}, \\ B^r(\mathbf{k}) &\equiv \{\theta \in M^r(\mathbf{k}) \cap (\hat{\Theta}(\mathbf{k}), \bar{\theta}]\}. \end{aligned}$$

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

Defining an alternative investment agreement $\hat{\mathbf{T}}$. Let the agreement $\hat{\mathbf{T}} = (\hat{T}^1, .., \hat{T}^h, ..., \hat{T}^H)$ be characterized by a threshold $\hat{\Theta}(\mathbf{k})$ given by

$$F(\hat{\Theta}(\mathbf{k})) \equiv \min\{\textstyle\int_{M(\mathbf{k})} dF(\theta); F(\Theta^J(\mathbf{k}))\} \tag{A.10}$$

and compensation payments for all firms $h \in \mathcal{H}$:

$$\hat{T}^h(\mathbf{k},\theta) = \begin{cases} \Pi^h(\mathbf{k}) & \theta \in A(\mathbf{k}) \cup A^r(\mathbf{k}) = [\underline{\theta}, \hat{\Theta}(\mathbf{k})] \\ \tilde{T}^h(\mathbf{k},\theta) & \theta \in B(\mathbf{k}) \\ T^h(\mathbf{k},\theta) & \theta \in B^r(\mathbf{k}) \end{cases}. \tag{A.11}$$

For $\int_{B(\mathbf{k})} dF(\tilde{\theta}) = 0$, let $\tilde{T}^h(\mathbf{k},\theta) = 0$. For $\int_{B(\mathbf{k})} dF(\tilde{\theta}) > 0$:

$$\tilde{T}^h(\mathbf{k},\theta) \equiv \frac{1}{\int_{B(\mathbf{k})} dF(\tilde{\theta})} [\textstyle\int_{A^r(\mathbf{k})} T^h(\mathbf{k},\tilde{\theta}) dF(\tilde{\theta}) + \max\{\int_{M(\mathbf{k})} dF(\tilde{\theta}) - F(\Theta^J(\mathbf{k})); 0\}\Pi^h(\mathbf{k})]. \tag{A.12}$$

Establishing $\hat{\Theta}(\mathbf{k}) \in [\Theta(\mathbf{k}), \Theta^J(\mathbf{k})]$. The inequality $\hat{\Theta}(\mathbf{k}) \leq \Theta^J(\mathbf{k})$ follows directly from (A.10). If $F(\hat{\Theta}(\mathbf{k})) = \int_{M(\mathbf{k})} dF(\theta)$, then $\hat{\Theta}(\mathbf{k}) \geq \Theta(\mathbf{k})$ by Lemma 1. If $F(\hat{\Theta}(\mathbf{k})) = F(\Theta^J(\mathbf{k}))$, then $\hat{\Theta}(\mathbf{k}) \geq \Theta(\mathbf{k})$ by $\Theta^J(\mathbf{k}) \geq \Theta(\mathbf{k})$.

The host country regulates under agreement $\hat{\mathbf{T}}$ if and only if $\theta > \hat{\Theta}(\mathbf{k})$. Consider the incentives for the host country to regulate the industry under an arbitrary investment profile $\mathbf{k}$ for agreement $\hat{\mathbf{T}}$ and for different realizations of the shock θ:
(i) $\theta \in A(\mathbf{k}) \cup A^r(\mathbf{k}) = [\underline{\theta}, \hat{\Theta}(\mathbf{k})]$. By construction of the agreement, the net benefit of allowing production is non-negative for all $\theta \leq \hat{\Theta}(\mathbf{k})$:

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

(ii) $\theta \in B^r(\mathbf{k})$. 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) $\theta \in B(\mathbf{k})$ and $\int_{B(\mathbf{k})} dF(\tilde{\theta}) = 0$. Firms receive zero compensation in this case, which implies

$$V(\mathbf{k},\theta) + \textstyle\sum_{h=1}^{H} T^h(\mathbf{k},\theta) = V(\mathbf{k},\theta) < V(\mathbf{k},\hat{\Theta}(\mathbf{k})) \leq V(\mathbf{k},\Theta(\mathbf{k})) \leq 0.$$

(iv) $\theta \in B(\mathbf{k})$ and $\int_{B(\mathbf{k})} dF(\tilde{\theta}) > 0$. By the construction of $\hat{\Theta}(\mathbf{k})$:

$$\textstyle\int_{B(\mathbf{k})} dF(\tilde{\theta}) \equiv \int_{A^r(\mathbf{k})} dF(\tilde{\theta}) + \max\{\int_{M(\mathbf{k})} dF(\tilde{\theta}) - F(\Theta^J(\mathbf{k})); 0\}. \tag{A.13}$$

Use $\tilde{T}^h(\mathbf{k},\theta)$ defined in (A.12), and (A.13) to decompose the net benefit of allowing production in the host country as follows:

$$\begin{aligned} & \textstyle\int_{B(\mathbf{k})} dF(\tilde{\theta})[V(\mathbf{k},\theta) + \sum_{h=1}^{H} \tilde{T}^h(\mathbf{k},\theta)] \\ = \; & \textstyle\int_{A^r(\mathbf{k})} [V(\mathbf{k},\theta) - V(\mathbf{k},\tilde{\theta})] dF(\tilde{\theta}) + \int_{A^r(\mathbf{k})} [V(\mathbf{k},\tilde{\theta}) + \sum_{h=1}^{H} T^h(\mathbf{k},\tilde{\theta})] dF(\tilde{\theta}) \\ & \textstyle + [V(\mathbf{k},\theta) - V(\mathbf{k},\Theta^J(\mathbf{k}))] \max\{\int_{M(\mathbf{k})} dF(\tilde{\theta}) - F(\Theta^J(\mathbf{k})); 0\}. \end{aligned}$$

Assume first that $\int_{A^r(\mathbf{k})} dF(\theta) > 0$. In this case, the first term on the second row is strictly negative because $V_\theta < 0$ and $\theta > \hat{\Theta}(\mathbf{k}) \geq \tilde{\theta}$ for all $\theta \in B(\mathbf{k})$ and $\tilde{\theta} \in A^r(\mathbf{k})$. The second term on the second row is strictly negative because regulation is optimal under contract $\mathbf{T}$ for all $\tilde{\theta} \in A^r(\mathbf{k})$. The term on the third row is zero if $\int_{M(\mathbf{k})} dF(\tilde{\theta}) \leq F(\Theta^J(\mathbf{k}))$ and strictly negative otherwise because then $\theta > \hat{\Theta}(\mathbf{k}) = \Theta^J(\mathbf{k})$ for all $\theta \in B(\mathbf{k})$. The terms on the second row vanish if $\int_{A^r(\mathbf{k})} dF(\theta) = 0$. But then $\int_{M(\mathbf{k})} dF(\theta) > F(\Theta^J(\mathbf{k}))$ by (A.13) and the assumption that $\int_{B(\mathbf{k})} dF(\tilde{\theta}) > 0$, so the third term is strictly negative in this case. We conclude that it is ex post strictly optimal for the host country to regulate if and only if $\theta > \hat{\Theta}(\mathbf{k})$ under the compensation rule $\hat{\mathbf{T}}$.

Investments and expected profits are the same under both agreements. By way of the threshold $\hat{\Theta}(\mathbf{k})$ for regulation defined in (A.10) and the compensation rules (A.11)-(A.12), the expected investment profit of every firm $h \in \mathcal{H}$ is the same under both compensation mechanisms for all $\mathbf{k}$:

$$\begin{aligned} & F(\hat{\Theta}(\mathbf{k}))\Pi^h(\mathbf{k}) + \tilde{T}^h(\mathbf{k},\theta) \textstyle\int_{B(\mathbf{k})} dF(\theta) + \int_{B^r(\mathbf{k})} T^h(\mathbf{k},\theta) dF(\theta) - R^h(k_h) \\ & \quad = \textstyle\int_{M(\mathbf{k})} dF(\theta)\Pi^h(\mathbf{k}) + \int_{M^r(\mathbf{k})} T^h(\mathbf{k},\theta) dF(\theta) - R^h(k_h). \end{aligned}$$

Hence, $\hat{\mathbf{k}}$ can be sustained as an equilibrium also under $\hat{\mathbf{T}}$. Furthermore, $\tilde{\Pi}^h(\hat{\mathbf{T}}) = \tilde{\Pi}^h(\mathbf{T})$ for all $h \in \mathcal{H}$, and therefore $\tilde{\Pi}(\hat{\mathbf{T}}) = \tilde{\Pi}(\mathbf{T})$.

Expected host country surplus is weakly higher under agreement $\hat{\mathbf{T}}$. The marginal utility of compensation payments is constant and the same for all parties. Therefore, $\tilde{V}(\hat{\mathbf{T}}) = \omega(\hat{\mathbf{T}}) - \tilde{\Pi}(\hat{\mathbf{T}})$ and $\tilde{V}(\mathbf{T}) = \omega(\mathbf{T}) - \tilde{\Pi}(\mathbf{T})$, where $\omega(\mathbf{T})$ denotes the expected joint surplus of general compensation $\mathbf{T}$; see (A.7). Hence,

$$\tilde{V}(\hat{\mathbf{T}}) - \tilde{V}(\mathbf{T}) = \omega(\hat{\mathbf{T}}) - \omega(\mathbf{T}) = \textstyle\int_{\underline{\theta}}^{\hat{\theta}} [V(\hat{\mathbf{k}},\theta) + \Pi(\hat{\mathbf{k}})] dF(\theta) - \int_{M(\hat{\mathbf{k}})} [V(\hat{\mathbf{k}},\theta) + \Pi(\hat{\mathbf{k}})] dF(\theta),$$

where $\hat{\theta} \equiv \hat{\Theta}(\hat{\mathbf{k}})$. Adding and subtracting $V(\hat{\mathbf{k}}, \hat{\theta})$ underneath the two integrals and rewriting yields

$$\begin{aligned}\omega(\hat{\mathbf{T}}) - \omega(\mathbf{T}) &= \textstyle\int_{A^r(\hat{\mathbf{k}})}[V(\hat{\mathbf{k}}, \theta) - V(\hat{\mathbf{k}}, \hat{\theta})]dF(\theta) + \int_{B(\hat{\mathbf{k}})}[V(\hat{\mathbf{k}}, \hat{\theta}) - V(\hat{\mathbf{k}}, \theta)]dF(\theta) \\ &\quad \textstyle +[V(\hat{\mathbf{k}}, \hat{\theta}) + \Pi(\hat{\mathbf{k}})][\min\{\int_{M(\hat{\mathbf{k}})} dF(\theta); F(\Theta^J(\hat{\mathbf{k}}))\} - \int_{M(\hat{\mathbf{k}})} dF(\theta)].\end{aligned}$$

The two expressions on the first row are both non-negative because $V(\hat{\mathbf{k}}, \theta)$ is decreasing in θ, $\theta \leq \hat{\theta}$ in the domain $A^r(\hat{\mathbf{k}})$, and $\theta > \hat{\theta}$ in the domain $B(\hat{\mathbf{k}})$. The term on the second row is obviously zero if $\int_{M(\hat{\mathbf{k}})} dF(\theta) \leq F(\Theta^J(\hat{\mathbf{k}}))$. It is zero also if $\int_{M(\hat{\mathbf{k}})} dF(\theta) > F(\Theta^J(\hat{\mathbf{k}}))$ because then $V(\hat{\mathbf{k}}, \hat{\theta}) + \Pi(\hat{\mathbf{k}}) = V(\hat{\mathbf{k}}, \Theta^J(\hat{\mathbf{k}})) + \Pi(\hat{\mathbf{k}}) = 0$. Hence, $\omega(\hat{\mathbf{T}}) \geq \omega(\mathbf{T})$, which concludes the proof.■

Remarks. Based on (A.11) and (A.12), we can write

$$\hat{T}^h(\mathbf{k}, \theta) \equiv \tilde{\Lambda}(\mathbf{k}, \theta)\Pi^h(\mathbf{k}) + \hat{\Lambda}(\mathbf{k}, \theta)T^h(\mathbf{k}, \theta) + \textstyle\int_{\underline{\theta}}^{\bar{\theta}} T^h(\mathbf{k}, \tilde{\theta})d\Lambda(\mathbf{k}, \tilde{\theta})$$

for almost all $\theta \in [\underline{\theta}, \bar{\theta}]$, where $\tilde{\Lambda}(\mathbf{k}, \theta) \geq 0$, $\hat{\Lambda}(\mathbf{k}, \theta) \geq 0$, $\Lambda(\mathbf{k}, \theta) \geq 0$ and

$$\tilde{\Lambda}(\mathbf{k}, \theta) + \hat{\Lambda}(\mathbf{k}, \theta) + \textstyle\int_{\underline{\theta}}^{\bar{\theta}} d\Lambda(\mathbf{k}, \tilde{\theta}) = 1.$$

Compensation for each firm under the alternative agreement $\hat{\mathbf{T}}$ therefore is a convex combination of operating profit and compensation under the original agreement $\mathbf{T}$, and where the weights depend on $(\mathbf{k}, \theta)$, but are the same for all H firms. This property implies that the modified scheme $\hat{\mathbf{T}}$ inherits a number of characteristics from the initial scheme $\mathbf{T}$. First, compensation is non-negative because operating profit is non-negative and the original compensation is non-negative ($\Pi^h \geq 0$ and $T^h \geq 0$ imply $\hat{T}^h \geq 0$). Second, it does not rely on excessive compensation (punitive damages) if this is not part of the original scheme ($T^h \leq \Pi^h$ implies $\hat{T}^h \leq \Pi^h$). 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 3 would thus hold also for stricter restrictions on compensation payments than non-negativity. It also shows that linear compensation rules that incorporate both operating profits and incurred capital costs are (weakly) superior to rules that compensate incurred capital costs only.

A.4 Proof of Proposition 4

The incentive compatibility conditions for $(\mathbf{k}^J, \theta^J)$ and $(\mathbf{k}^{*J}, \theta^{*J})$ are $F(\theta^{JC}) \in [\max\{\Xi(\mathbf{k}^J); F(\theta^J)\}, 1]$ and $F^*(\theta^{*JC}) \in [\max\{\Xi^*(\mathbf{k}^{*J}); F^*(\theta^{*J})\}, 1]$ by the assumption of separability between the two countries. Hence, implementation of jointly efficient outcomes is neither easier or more difficult than under one-way flows. Consider a jointly negotiated two-way carve-out compensation agreement

$(\mathbf{T}^C, \mathbf{T}^{C*})$ with investment protection $(\theta^{JC}, \theta^{*JC})$ characterized by

$$F(\theta^{JC})\Pi(\mathbf{k}^J) - F^*(\theta^{*JC})\Pi^*(\mathbf{k}^{*J}) = (1-\alpha)(\omega^J - \omega^0) + \pi^0 + R(\mathbf{k}^J) - \alpha(\omega^{*J} - \omega^{*0}) - \pi^{*0} - R^*(\mathbf{k}^{*J}).$$

Plugging $(\theta^{JC}, \theta^{*JC})$ into the Nash product $\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*)$ of Section 4.4 yields

$$\mathcal{N}^B(\mathbf{T}^C, \mathbf{T}^{C*}) \equiv \alpha^\alpha(1-\alpha)^{1-\alpha}(\omega^J - \omega^0 + \omega^{*J} - \omega^{*0}).$$

Instead, the unconstrained Nash product under two-way investment flows equals

$$\hat{\mathcal{N}}^B(\mathbf{T}, \mathbf{T}^*, s) \equiv [\tilde{V}(\mathbf{T}) + \tilde{\Pi}^*(\mathbf{T}^*) - s - v^0 - \pi^{*0}]^\alpha[\tilde{V}^*(\mathbf{T}^*) + \tilde{\Pi}(\mathbf{T}) + s - v^{*0} - \pi^0]^{1-\alpha}$$

under an agreement with general compensation schemes $(\mathbf{T}, \mathbf{T}^*)$ and with unlimited side payments. The side-payment $S^B(\mathbf{T}, \mathbf{T}^*)$ that maximizes $\hat{\mathcal{N}}^B(\mathbf{T}, \mathbf{T}^*, s)$ yields

$$\hat{\mathcal{N}}^B(\mathbf{T}, \mathbf{T}^*, S^B(\mathbf{T}, \mathbf{T}^*)) = \alpha^\alpha(1-\alpha)^{1-\alpha}(\omega(\mathbf{T}) - \omega^0 + \omega^*(\mathbf{T}^*) - \omega^{*0}) \geq \hat{\mathcal{N}}^B(\mathbf{T}, \mathbf{T}^*, 0) = \mathcal{N}^B(\mathbf{T}, \mathbf{T}^*).$$

$\mathcal{N}^B(\mathbf{T}^C, \mathbf{T}^{C*}) \geq \mathcal{N}^B(\mathbf{T}, \mathbf{T}^*)$ then follows from $\omega^J \geq \omega(\mathbf{T})$ and $\omega^{*J} \geq \omega^*(\mathbf{T}^*)$. Investment protection θ^{JC} in Home, given by (10), and θ^{*JC} in Foreign, given by

$$F^*(\theta^{*JC}) = \frac{R^*(\mathbf{k}^{*J}) + \pi^{*0}}{\Pi^*(\mathbf{k}^{*J})} + \alpha\frac{\omega^{*J} - \omega^{*0}}{\Pi^*(\mathbf{k}^{*J})}, \tag{A.14}$$

under separate negotiations over one-ways flows also maximize the Nash Product $\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*)$ and therefore are optimal in the present setting. But the agreement with two-way investment flows can potentially do more. Suppose that $\Xi^*(\mathbf{k}^{*J}) \leq F^*(\theta^{*J})$ and $\theta^{*JC} = \theta^{*J} - \varepsilon$ in (A.14). Then no $\mathbf{T}^{*C}$ can implement $(\mathbf{k}^{*J}, \theta^{*J})$ and maximize $\mathcal{N}(\mathbf{T}^*)$ under one-way investment flows. Assume, however, that θ^{JC} in (10) satisfies $\max\{\Xi(\mathbf{k}^J); F(\theta^J)\} < F(\theta^{JC}) < 1$. Setting $\theta^{*JC} = \theta^{*J}$ and increasing θ^{JC} achieves the desired distribution of surplus under two-way investment flows. The additional flexibility in distributing investment protection across countries under joint agreement over two-way investment flows makes it easier to negotiate an efficient agreement.

To see that $\theta^{JC} = \theta^{*JC}$ sometimes is feasible also under asymmetries, start out with symmetry in both countries and assume that $\max\{\Xi(\mathbf{k}^J); F(\theta^J)\} < F(\theta^{JC}) < 1$. Introduce a small asymmetry, so that $F(\theta^{JC}) \neq F^*(\theta^{*JC})$ if defined by (10) and (A.14). In this case, $(\mathbf{T}^C, \mathbf{T}^{*C})$ implements $(\mathbf{k}^J, \theta^J, \mathbf{k}^{*J}, \theta^{*J})$ and maximizes $\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*)$. However, this is true also for $\theta^{JC} = \theta^{*JC} = \xi^C$ defined by

$$\frac{F(\xi^C) - F(\theta^{JC})}{F^*(\xi^C) - F^*(\theta^{*JC})} = \frac{\Pi^*(\mathbf{k}^{*J})}{\Pi(\mathbf{k}^J)}$$

ξ^C is feasible if asymmetry is sufficiently small because then ξ^C is similar both to θ^{JC} and θ^{*JC}.■

A.5 Proof of Proposition 7

Consider a carve-out mechanism $\mathbf{T}^C \equiv (\mathbf{T}^{rC}, \mathbf{T}^{xC})$. For any investment portfolio $\mathbf{k}$, the sequentially optimal host country decision can be partitioned into three subsets depending on the shock θ: The subset $M(\mathbf{k})$ for which non-intervention is ex post optimal, the subset $M^r(\mathbf{k})$ for which regulation of the industry is optimal, and the subset $M^x(\mathbf{k})$ for which the host country seizes the assets of the industry and continues operations. The profit maximizing investment of firm $h \in \mathcal{H}$ satisfies

$$\hat{k}^h \in \arg\max_{k_h \geq 0} \int_{M(k_h, \hat{\mathbf{k}}_{-h})} \Pi^h(k_h, \hat{\mathbf{k}}_{-h}) dF(\theta) + \int_{M^r(k_h, \hat{\mathbf{k}}_{-h})} \beta(k_h, \hat{\mathbf{k}}_{-h}, \theta) \Pi^h(k_h, \hat{\mathbf{k}}_{-h}) dF(\theta) - R^h(k_h),$$

where $\beta(\mathbf{k}, \theta) \in \{0, 1\}$ because $\mathbf{T}^C$ is a carve-out mechanism. Observe that the expected investment profit does not depend directly on $M^x(\mathbf{k})$ because the host country only expropriates the industry's assets if it does not have to pay any compensation for doing so. Summing up over all firms yields the equilibrium expected investment profit

$$\tilde{\Pi}(\mathbf{T}^C) = \int_{M(\hat{\mathbf{k}})} \Pi(\hat{\mathbf{k}}) dF(\theta) + \int_{M^r(\hat{\mathbf{k}})} \beta(\hat{\mathbf{k}}, \theta) \Pi(\hat{\mathbf{k}}) dF(\theta) - R(\hat{\mathbf{k}})$$

under $\mathbf{T}^C$, whereas the expected host country surplus can be written as

$$\tilde{V}(\mathbf{T}^C) \equiv \int_{M(\hat{\mathbf{k}})} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}})) dF(\theta) + \int_{M^x(\hat{\mathbf{k}})} (V(\hat{\mathbf{k}}, \theta) + \Pi^x(\hat{\mathbf{k}})) dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^C).$$

In the proof of this proposition we will rely on two general lemmas, the first of which is quite trivial:

Lemma 4 *For any initial carve-out mechanism $\mathbf{T}^C \equiv (\mathbf{T}^{rC}, \mathbf{T}^{xC})$ for which there is no direct expropriation in equilibrium, there exists an alternative carve-out mechanism $\mathbf{T}^{C\prime} \equiv (\mathbf{T}^{rC\prime}, \mathbf{T}^{xC\prime})$ that:*
(i) features complete investment protection from direct expropriation, $\Theta^{rC\prime}(\mathbf{k}) = \bar{\theta}$ for all $\mathbf{k}$; and
(ii) yields the same expected operating profit and host country expected surplus compared to the initial mechanism ($\tilde{\Pi}^h(\mathbf{T}^{C\prime}) = \tilde{\Pi}^h(\mathbf{T}^C)$ for all $h \in \mathcal{H}$, and $\tilde{V}(\mathbf{T}^{C\prime}) = \tilde{V}(\mathbf{T}^C)$).

Proof: In the alternative mechanism $\mathbf{T}^{C\prime} \equiv (\mathbf{T}^{rC\prime}, \mathbf{T}^{xC\prime})$ with $\Theta^{xC\prime}(\mathbf{k}) = \bar{\theta}$, let $\Theta^{rC'}(\mathbf{k})$ be defined by

$$F(\Theta^{rC'}(\mathbf{k})) \equiv \int_{M(\mathbf{k})} dF(\theta) + \int_{M^r(\mathbf{k})} \beta(\mathbf{k}, \theta) dF(\theta) \leq 1.$$

By $\Theta^{xC\prime}(\mathbf{k}) = \bar{\theta}$, the host country never seizes any assets. As $\Theta^{rC'}(\mathbf{k}) \geq \Theta(\mathbf{k})$ by $[\underline{\theta}, \Theta(\mathbf{k})] \subset M(\mathbf{k})$, see Lemma 1, the host country regulates for all $\theta > \Theta^{rC'}(\mathbf{k})$. The expected profit of firm $h \in \mathcal{H}$ of investing k_h thus equals

$$F(\Theta^{rC'}(k_h, \hat{\mathbf{k}}_{-h})) \Pi^h(k_h, \hat{\mathbf{k}}_{-h}) - R^h(k_h)$$

if all other firms invest $\hat{\mathbf{k}}_{-h}$; see Lemma 2. By construction of $\Theta^{rC'}(\mathbf{k})$, $\hat{k}_h$ is an optimal investment for h. Hence, $\hat{\mathbf{k}}$ can be sustained as an equilibrium investment portfolio under $\mathbf{T}^{C\prime}$. Moreover, all firms have the same expected investment profit as in the initial mechanism, which implies

$$\tilde{\Pi}(\mathbf{T}^{C'}) = F(\theta^{rC'})\Pi(\hat{\mathbf{k}}) - R(\hat{\mathbf{k}}) = \tilde{\Pi}(\mathbf{T}^{C}),$$

where $\hat{\theta}^{rC'} \equiv \Theta^{rC'}(\hat{\mathbf{k}})$. The host country's expected surplus equals

$$\tilde{V}(\mathbf{T}^{C'}) = \tilde{V}(\mathbf{T}^{C}) = \int_{\underline{\theta}}^{\min\{\hat{\theta}^{rC'};\hat{\theta}^{J}\}} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^{C}),$$

which completes the proof.■

Of course, there are many mechanisms that yield direct expropriation in equilibrium. More generally:

Lemma 5 *For any initial carve-out mechanism $\mathbf{T}^{C} \equiv (\mathbf{T}^{rC}, \mathbf{T}^{xC})$, there exists an alternative carve-out mechanism $\mathbf{T}^{C\prime} \equiv (\mathbf{T}^{rC\prime}, \mathbf{T}^{xC\prime})$ that:*
(i) features symmetric investment protection $\Theta^{rC\prime}(\mathbf{k}) = \Theta^{xC\prime}(\mathbf{k})$ for all $\mathbf{k}$; and
(ii) yields the same expected operating profit and weakly higher host country expected surplus compared to the initial mechanism ($\tilde{\Pi}^h(\mathbf{T}^{C\prime}) = \tilde{\Pi}^h(\mathbf{T}^{C})$ for all $h \in \mathcal{H}$, and $\tilde{V}(\mathbf{T}^{C\prime}) \geq \tilde{V}(\mathbf{T}^{C})$).

Proof: In the alternative mechanism $\mathbf{T}^{C\prime} \equiv (\mathbf{T}^{rC\prime}, \mathbf{T}^{xC\prime})$, let $\Theta^{rC'}(\mathbf{k}) = \Theta^{xC'}(\mathbf{k}) = \Theta^{C'}(\mathbf{k})$, where

$$F(\Theta^{C'}(\mathbf{k})) \equiv \int_{M(\mathbf{k})} dF(\theta) + \int_{M^r(\mathbf{k})} \beta(\mathbf{k}, \theta)dF(\theta) \leq 1.$$

As $\Theta^{C'}(\mathbf{k}) \geq \Theta(\mathbf{k})$ by $[\underline{\theta}, \Theta(\mathbf{k})] \subset M(\mathbf{k})$, see Lemma 1, the host country intervenes for all $\theta > \Theta^{C'}(\mathbf{k})$ either by regulation or by seizing the industry's assets, and without paying compensation either way. The expected profit of firm $h \in \mathcal{H}$ of investing k_h thus equals

$$F(\Theta^{C'}(k_h, \hat{\mathbf{k}}_{-h}))\Pi^h(k_h, \hat{\mathbf{k}}_{-h}) - R^h(k_h)$$

if all other firms invest $\hat{\mathbf{k}}_{-h}$. By construction of $\Theta^{C'}(\mathbf{k})$, $\hat{k}_h$ is an optimal investment for h. Hence, $\hat{\mathbf{k}}$ can be sustained as an equilibrium investment portfolio under $\mathbf{T}^{C\prime}$. Moreover, all firms have the same expected investment profit as in the initial mechanism, which implies

$$\tilde{\Pi}(\mathbf{T}^{C'}) = F(\hat{\theta}^{C'})\Pi(\hat{\mathbf{k}}) - R(\hat{\mathbf{k}}) = \tilde{\Pi}(\mathbf{T}^{C}),$$

where $\hat{\theta}^{C'} \equiv \Theta^{C'}(\hat{\mathbf{k}})$. If we let $\hat{\theta}' \equiv \min\{\hat{\theta}^{C}; \hat{\theta}^{J}\}$, then the expected host country surplus under $\mathbf{T}^{C\prime}$ equals

$$\tilde{V}(\mathbf{T}^{C\prime}) \equiv \int_{\underline{\theta}}^{\hat{\theta}'} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) + \int_{\hat{\theta}'}^{\max\{\hat{\theta}'; \hat{\theta}^{x}\}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^{C'})$$

We proceed case-by-case to establish $\tilde{V}(\mathbf{T}^{C\prime}) \geq \tilde{V}(\mathbf{T}^{C})$.

Case (i): $\hat{\theta}^{C'} \geq \hat{\theta}^{J}$. We can write the difference in expected host country surplus under the two mechanisms as

$$\begin{aligned}
\tilde{V}(\mathbf{T}^{C\prime}) - \tilde{V}(\mathbf{T}^{C}) \quad = \quad & \int_{M^{r}(\hat{\mathbf{k}}) \cap [\underline{\theta}, \hat{\theta}^{J}]} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) + \int_{M^{x}(\hat{\mathbf{k}}) \cap [\underline{\theta}, \hat{\theta}^{J}]} (\Pi(\hat{\mathbf{k}}) - \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) \\
& - \int_{M(\hat{\mathbf{k}}) \cap (\hat{\theta}^{J}, \bar{\theta}]} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) - \int_{M^{x}(\hat{\mathbf{k}}) \cap (\hat{\theta}^{J}, \bar{\theta}]} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta)
\end{aligned}$$

The first term on the first row is non-negative because $V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}) \geq 0$ for all $\theta \leq \hat{\theta}^{J}$. The second term on the first row is non-negative by $\Pi(\hat{\mathbf{k}}) \geq \Pi^{x}(\hat{\mathbf{k}})$. The two terms on the second row are non-negative by $V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}) \leq V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}) < 0$ for all $\theta > \hat{\theta}^{J}$.

Case (ii): $\hat{\theta}^{rC} \leq \hat{\theta}^{J}$. In this case, the firm will optimally allow production for all $\theta \leq \hat{\theta} \equiv \min\{\hat{\theta}^{rC}; \hat{\theta}^{xC}\}$ and intervene for all $\theta > \hat{\theta}$ under the initial contract. The host country expected utility equals:

$$\tilde{V}(\mathbf{T}^{C}) = \int_{\underline{\theta}}^{\hat{\theta}} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) + \int_{\hat{\theta}^{xC}}^{\max\{\hat{\theta}^{xC}; \hat{\theta}^{x}\}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^{C}).$$

Looking next at $\tilde{V}(\mathbf{T}^{C\prime})$, $\hat{\theta}^{rC} \leq \hat{\theta}^{J}$ implies that there will never be any compensation payments if the host country regulates. Hence, $F(\hat{\theta}^{C'}) = \int_{M(\hat{\mathbf{k}})} dF(\theta) = F(\hat{\theta})$ and therefore

$$\tilde{V}(\mathbf{T}^{C\prime}) = \int_{\underline{\theta}}^{\hat{\theta}} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) + \int_{\hat{\theta}}^{\max\{\hat{\theta}; \hat{\theta}^{x}\}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^{C'}).$$

Subtracting the two yields

$$\tilde{V}(\mathbf{T}^{C\prime}) - \tilde{V}(\mathbf{T}^{C}) = \int_{\hat{\theta}}^{\max\{\hat{\theta}; \hat{\theta}^{x}\}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) - \int_{\hat{\theta}^{xC}}^{\max\{\hat{\theta}^{xC}; \hat{\theta}^{x}\}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta).$$

The difference is zero if $\hat{\theta}^{xC} \leq \hat{\theta}^{rC}$ because then $\hat{\theta} = \hat{\theta}^{xC}$. It is zero also if $\hat{\theta}^{x} \leq \hat{\theta}^{rC} < \hat{\theta}^{xC}$, because then $\max\{\hat{\theta}; \hat{\theta}^{x}\} = \hat{\theta}$ and $\max\{\hat{\theta}^{xC}; \hat{\theta}^{x}\} = \hat{\theta}^{xC}$. If $\hat{\theta}^{rC} < \hat{\theta}^{xC}$ and $\hat{\theta}^{x} \in (\hat{\theta}^{rC}, \hat{\theta}^{xC}]$, then

$$\tilde{V}(\mathbf{T}^{C\prime}) - \tilde{V}(\mathbf{T}^{C}) = \int_{\hat{\theta}^{rC}}^{\hat{\theta}^{x}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) > 0.$$

If $\hat{\theta}^{rC} < \hat{\theta}^{xC} < \hat{\theta}^{x}$, then

$$\tilde{V}(\mathbf{T}^{C\prime}) - \tilde{V}(\mathbf{T}^{C}) = \int_{\hat{\theta}^{rC}}^{\hat{\theta}^{xC}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) > 0.$$

Case (iii): $\hat{\theta}^{rC} > \hat{\theta}^{J}$ and $\hat{\theta}^{xC} \geq \hat{\theta}^{J}$. In this case, the firm is allowed to produce for all $\theta \leq \hat{\theta}^{J}$ under the initial agreement $\mathbf{T}^{C}$, whereas there is intervention for all $\theta > \hat{\theta}^{J}$. Hence,

$$F(\hat{\theta}^{C'}) = F(\hat{\theta}^{J}) + \int_{\hat{\theta}^{J}}^{\bar{\theta}} \beta(\hat{\mathbf{k}}, \theta)dF(\theta) \geq F(\hat{\theta}^{J})$$

and therefore Case (i) applies.

Case (iv): $\hat{\theta}^{rC} \in (\hat{\theta}^{J}, \hat{\theta}^{Y}]$ and $\hat{\theta}^{xC} < \hat{\theta}^{J}$, where $V(\hat{\mathbf{k}}, \hat{\theta}^{Y}) + \Pi(\hat{\mathbf{k}}) + \Pi^{x}(\hat{\mathbf{k}}) \equiv 0$ if $V(\hat{\mathbf{k}}, \bar{\theta}) + \Pi(\hat{\mathbf{k}}) + \Pi^{x}(\hat{\mathbf{k}}) < 0$ and $\hat{\theta}^{Y} = \bar{\theta}$ otherwise. In this case, the host country allows production for all $\theta \leq \hat{\theta}^{xC}$, seizes the industry's assets for all $\theta \in (\hat{\theta}^{xC}, \hat{\theta}^{rC}]$, and regulates for all $\theta > \hat{\theta}^{rC}$. Under the initial contract, therefore,

$$\tilde{V}(\mathbf{T}^{C}) = \int_{\underline{\theta}}^{\hat{\theta}^{xC}} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) + \int_{\hat{\theta}^{xC}}^{\hat{\theta}^{rC}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^{C}).$$

Under the initial mechanism $\mathbf{T}^{C}$, the host country regulates the industry if and only if $\theta > \hat{\theta}^{rC}$, in which case it does not have to pay compensation. Hence, $F(\hat{\theta}^{C'}) = \int_{M(\hat{\mathbf{k}})} dF(\theta) = F(\hat{\theta}^{xC})$, and expected host country surplus therefore equals

$$\tilde{V}(\mathbf{T}^{C\prime}) = \int_{\underline{\theta}}^{\hat{\theta}^{xC}} (V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) + \int_{\hat{\theta}^{xC}}^{\max\{\hat{\theta}^{xC}; \theta^{x}\}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^{C\prime}).$$

The difference in surplus is positive

$$\tilde{V}(\mathbf{T}^{C\prime}) - \tilde{V}(\mathbf{T}^{C}) = -\int_{\max\{\hat{\theta}^{xC}; \theta^{x}\}}^{\hat{\theta}^{rC}} (V(\hat{\mathbf{k}}, \theta) + \Pi^{x}(\hat{\mathbf{k}}))dF(\theta)$$

because it is better to shut down production than to expropriate the assets and continue production for all $\theta > \theta^{x}$. If $\hat{\theta}^{Y} = \bar{\theta}$, the proof is done. Consider therefore the final case of $\theta^{Y} < \bar{\theta}$.

Case (v): $\hat{\theta}^{rC} > \hat{\theta}^{Y}$ and $\hat{\theta}^{xC} < \hat{\theta}^{J}$. The host country allows production for all $\theta \leq \hat{\theta}^{xC}$, seizes the industry's assets for all $\theta \in (\hat{\theta}^{xC}, \hat{\theta}^{Y})$, regulates and pays compensation for all $\theta \in [\hat{\theta}^{Y}, \hat{\theta}^{rC}]$, and regulates without paying compensation for all $\theta > \hat{\theta}^{rC}$ under $\mathbf{T}^C$. If $F(\hat{\theta}^{C'}) = F(\hat{\theta}^{xC}) + F(\hat{\theta}^{rC}) - F(\hat{\theta}^{Y}) \geq F(\hat{\theta}^{J})$, then Case (i) applies and we are done. Assume therefore that $F(\hat{\theta}^{C'}) < F(\hat{\theta}^{J})$. Under the initial mechanism

$$\tilde{V}(\mathbf{T}^C) = \int_{\underline{\theta}}^{\hat{\theta}^{xC}} (V(\hat{\mathbf{k}},\theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) + \int_{\hat{\theta}^{xC}}^{\hat{\theta}^{Y}} (V(\hat{\mathbf{k}},\theta) + \Pi^x(\hat{\mathbf{k}}))dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^C),$$

whereas

$$\tilde{V}(\mathbf{T}^{C\prime}) = \int_{\underline{\theta}}^{\hat{\theta}^{C'}} (V(\hat{\mathbf{k}},\theta) + \Pi(\hat{\mathbf{k}}))dF(\theta) + \int_{\hat{\theta}^{C'}}^{\max\{\hat{\theta}^{C};\theta^x\}} (V(\hat{\mathbf{k}},\theta) + \Pi^x(\hat{\mathbf{k}}))dF(\theta) - R(\hat{\mathbf{k}}) - \tilde{\Pi}(\mathbf{T}^{C\prime})$$

under the modified mechanism. Subtracting the two yields

$$\tilde{V}(\mathbf{T}^{C\prime}) - \tilde{V}(\mathbf{T}^C) \equiv \int_{\hat{\theta}^{xC}}^{\hat{\theta}^{C}} (\Pi(\hat{\mathbf{k}}) - \Pi^x(\hat{\mathbf{k}}))dF(\theta) - \int_{\max\{\hat{\theta}^{C'};\theta^x\}}^{\hat{\theta}^{Y}} (V(\hat{\mathbf{k}},\theta) + \Pi^x(\hat{\mathbf{k}}))dF(\theta),$$

which is strictly positive because $\Pi^x(\hat{\mathbf{k}}) \leq \Pi(\hat{\mathbf{k}})$ and $V(\hat{\mathbf{k}},\theta) + \Pi^x(\hat{\mathbf{k}}) < 0$ for all $\theta > \theta^x$.■

Now to the main results. Part 2 of Proposition 7 follows directly from Lemma 5 because the host country and investors always weakly benefit from replacing an initial carve-out mechanism $\mathbf{T}^C = (\mathbf{T}^{rC}, \mathbf{T}^{xC})$ by an alternative carve-out mechanism $\mathbf{T}^{C\prime} \equiv (\mathbf{T}^{rC\prime}, \mathbf{T}^{xC\prime})$ that features $\Theta^{rC\prime}(\mathbf{k}) = \Theta^{xC\prime}(\mathbf{k})$ for all $\mathbf{k}$. This holds for any combination of $\hat{\theta}^{rC}$, $\hat{\theta}^{x}$ and $\hat{\theta}^{J}$, not only those with $\hat{\theta}^{rC} < \hat{\theta}^{x} \leq \hat{\theta}^{J}$. To prove Part 1 of Proposition 7 consider without loss of generality an initial mechanism with $\hat{\theta}^{rC} = \hat{\theta}^{xC}$. If either $\hat{\theta}^{x} \leq \hat{\theta}^{J} \leq \hat{\theta}^{rC}$ or $\hat{\theta}^{x} \leq \hat{\theta}^{rC} < \hat{\theta}^{J}$ there will never be expropriation in equilibrium under the initial mechanism $\mathbf{T}^C = (\mathbf{T}^{rC}, \mathbf{T}^{xC})$. Lemma 4 then applies.■