Journal of International Economics
Peer-reviewed and accepted version

Economics of International Investment Agreements

Henrik Horn, Thomas Tangerås

Published version:
https://doi.org/10.1016/j.jinteco.2021.103433

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

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

December 30, 2020

[1]We are grateful to John Bonin, Robert L. Howse, Petros C. Mavroidis, Damien J. Neven, and two anonymous reviewers 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 2 700 highly potent international investment agreements protect foreign investment against host country policies. This paper analyzes the design and implications of their contentious provisions regarding regulatory expropriations. It derives conditions under which "carve-out" compensation mechanisms, similar to those in actual agreements, solve underinvestment and overregulation problems and simultaneously distribute surplus according to countries' bargaining power. The paper examines a number of additional policy relevant issues, for instance, how to modify agreements when carve-out compensation is inefficient, whether agreements cause "regulatory chill," and the different motives and distributional consequences of North-South versus North-North agreements.

JEL Codes: F21; F23; F53; K33

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

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 include 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 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? Who benefits and who loses? Why are some agreements more controversial than others? Do agreements between poor and rich countries serve the same purposes as those 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, 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. This agreement specifies how much compensation investors shall receive for regulation in different circumstances.[4] When negotiating the agreement, the source

[4]The agreement cannot simply prohibit regulation since regulation is jointly desirable for severe regulatory shocks.

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 to focus on contract stipulations that are consistent with those found in actual agreements, to highlight the efficiency properties and consequences of these agreements, and to explore 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 across agreements in the protection they 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 they receive no compensation if they are 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 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 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

This property renders the setting qualitatively different from a standard hold-up problem in which intervention can only reduce joint surplus.

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 even asymmetric countries may negotiate a reciprocal agreement in which the same conditions apply to both signatory countries. This finding sheds light on the reciprocal nature of actual investment agreements.

Our final set of findings concern 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 stimulate investment precisely by reducing regulation. We also establish that equilibrium investment agreements will not yield underregulation from a joint perspective under standard circumstances.

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 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 expected surplus in South since 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 parties 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 the expected surplus generated domestically 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, the countries will negotiate more investment protection than what maximizes expected domestic surplus. We thus find that the North-North agreement benefits investors at the expense of the rest of society. Lacking a general equilibrium representation of the economies involved, we cannot say anything definitive about distributional impact of a North-North agreement. But the finding that only foreign investors benefit from the direct effects of an agreement in the present setup, might explain some of 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 instead face the problem 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, 2018, 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) considers distributional effects of an exogenously specified agreement, assuming that firms anticipate the effects of investment on regulation. Increased investment protection from the agreement benefits investors, but exacerbates an already existing overinvestment and underregulation. The host country is worse off and total surplus can also go down compared to the situation without an agreement. 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

[5] An 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 focus 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 this literature and discusses a range of issues related to insecurity of property rights and foreign investment.

compensation when regulatory policies are changed in unfavorable direction for investors, with an agreement that instead comprises a National Treatment provision. In Schjelderup and Stähler (2020), firms overinvest because of market power. It is shown that an exogenously imposed agreement can reduce host country and global welfare. 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 exogenously specified agreements on the parties. In particular, 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.[6]

Closer to this paper is a second line of investigation that analyzes the design of efficient agreements. 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.[7] 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 (2020) 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. Focusing on direct expropriation and dispute settlement, they address issues that are complementary to those studied here.[8]

Our analysis focuses on endogenously and purposefully designed agreements, similar to the four papers just mentioned. But in contrast to those papers, and the rest of the literature, we develop

[6] In a discussion paper version (Horn and Tangerås, 2017), we extend the analysis of our baseline model by considering two other standard provisions in investment agreements. We show that negotiated compensation schemes for direct expropriation either entail complete investment protection against direct expropriation, or offer the same investment protection for all types of expropriation. The paper also establishes that agreements that only feature National Treatment are Pareto dominated by those with carve-out compensation.

[7] 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.

[8] Horn and Tangerås (2020) apply the framework developed in the current paper to examine differences between state-state and investor-state dispute settlement.

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- and two-way investment flows, and in examining how countries' abilities to make unilateral commitments affect the negotiated outcomes. 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 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 discusses policy issues. Section 6 concludes and contains suggestions for future research. All formal proofs appear in the Appendix, which also elaborates on some extensions of the baseline model.

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.[9] There is a single industry with $H \geq 1$ Foreign risk-neutral firms.[10] 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 a health hazard. To represent this feature, we assume that an industry-specific shock θ affecting the surplus 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

[9] We consider two-way investment in the Appendix to Section 4.3.

[10] A defining characteristic of international investment agreements is their economy-wide scope. We extend the model to cover multiple industries in the Appendix to Section 4.3. This extension has some interesting implications regarding efficiency, but does not fundamentally change the analysis.

"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.[11] Let $V(\mathbf{k}, \theta)$ denote the host country surplus 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 effect of investment by firm h on the surplus can be positive or negative, $V_h(\mathbf{k}, \theta) \gtrless 0$ (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 surplus from 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 surplus is non-negative at the most favorable realization of the shock, $V(\mathbf{k}, \underline{\theta}) \geq 0$, and 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 surplus $V(\mathbf{k}, \theta)$. In general, it could be in the host country's interest to regulate only a subset of firms. We take our definition of an industry to mean a set of firms that contribute in a sufficiently similar manner to Home's surplus 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 surplus is zero, and all H firms receive zero operating profits.

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 surplus $V(\mathbf{k}, \theta)$ 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 $\mathbf{k}$, 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.

[11] The Appendix to Section 4.3 analyzes partial regulation.

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 invests

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

in subgame-perfect equilibrium, where $\mathbf{k}_{-h}^0 = (k_1^0, ..., k_{h-1}^0, k_{h+1}^0, ..., k_H^0)$ constitutes the equilibrium investment portfolio of all firms other than h.

Let v^0 be the expected host country surplus and π^0 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. We let $\omega^0 \equiv v^0 + \pi^0$ denote the joint expected surplus of the host country and foreign investors absent any agreement

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 efficient 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 and 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 jointly efficient regulation, the expected joint surplus of the host country and foreign firms equals

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

The jointly efficient 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 jointly efficient threshold for regulation evaluated at the jointly efficient 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. We let $\omega^J \equiv \Omega(\mathbf{k}^J)$ be the joint expected surplus at the jointly efficient outcome $(\mathbf{k}^J, \theta^J)$.

There are several reasons why the equilibrium outcome $(\mathbf{k}^0, \theta^0)$ would generally differ from the jointly efficient outcome $(\mathbf{k}^J, \theta^J)$ and therefore generate inferior surplus $\omega^0 < \omega^J$. 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. 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.[12]

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

[12] 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.

portfolio $\mathbf{k}$ and the magnitude of the regulatory shock θ completely describe the host country surplus and investor profits, the compensation function $T^h = T^h(\mathbf{k}, \theta)$ for each firm h embodies 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.
3. The regulatory shock θ is observed; and
4. Home chooses whether to:
   (a) Permit production, in which case Home surplus is $V(\mathbf{k}, \theta)$ and each firm h receives $\Pi^h(\mathbf{k})$, or:
   (b) Regulate and pay compensation, in which case Home surplus 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.[13] 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 paid without formal arbitration.[14]

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. Due to perfect contract enforcement, it would always be better for the host country to pay compensation compared to re-allowing production; we will return to this below.

[13] 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.

[14] Fundamental results of our paper do not depend crucially on such perfect enforcement. We discuss implications of imperfect enforcement in the Appendix to Section 4.3.

3.2 Compensation schemes

We want to restrict the structure of the agreement under consideration to reflect compensation mechanisms in actual agreements, although we will also examine the extent to which these features constrain the outcomes. 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:

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...."[15]

[15] 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).

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 regulatory 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 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, i.e. in excess of foregone operating profits. This important principle in international law is increasingly often explicitly stated in investment agreements.[16] We discuss 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.[17] 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.[18]

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

[17] The Appendix to Section 4.3 demonstrates additional benefits of writing an economy-wide agreement.

[18] 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 (2020) refer to as "cease and

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

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

for each firm h given the equilibrium investment profile $\hat{\mathbf{k}}_{-h} = (\hat{k}_1, ..., \hat{k}_{h-1}, \hat{k}_{h+1}, ..., \hat{k}_H)$ 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.

The expected surplus of Home equals

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

under an agreement based on general compensation $\mathbf{T}$. We assume that Foreign's expected surplus 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

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

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 $V(\mathbf{k}^0, \theta) < 0$). 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 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.

The expected joint surplus of Home and Foreign under $\mathbf{T}$ is

$$\omega(\mathbf{T}) \equiv \tilde{V}(\mathbf{T}) + \tilde{\Pi}(\mathbf{T}) = \int_{M(\hat{\mathbf{k}})} [V(\hat{\mathbf{k}}, \theta) + \Pi(\hat{\mathbf{k}})]dF(\theta) - R(\hat{\mathbf{k}}) \quad (6)$$

All transfer payments cancel out because the marginal effect of compensation payments on the surplus is constant and the same for the two contracting parties. $\mathbf{T}$ has only indirect effects on joint surplus through the effect on regulation and investment. By implication, any agreement that implements the jointly efficient outcome ($\mathbf{k}^J$,θ^J) also maximizes the expected joint surplus $\omega(\mathbf{T})$.

desist," yields host country surplus $V(\mathbf{k}, \theta)$, 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 $M^r(\mathbf{k})$. Hence, cease and desist can be optimal only if contract enforcement is imperfect. See Ossa, Staiger and Sykes (2020) for a detailed analysis.

4 Properties of negotiated regulatory expropriation provisions

Consider the negotiations over an agreement in the setting laid out in Sections 2 and 3. 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. 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.

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. In Section 4.1, we analyze agreements that implement efficient outcomes. We consider agreements with inefficient outcomes in Section 4.2. Section 4.3 extends the benchmark analysis.

4.1 When carve-out compensation can implement the jointly efficient outcome

The potential of the contracting parties to negotiate jointly efficient outcomes depends on the extent to which parties can use side payments. We consider the outcomes under the two polar assumptions of unconstrained side payments versus no side payments.

4.1.1 Unconstrained side payments

Investment protection provisions are sometimes negotiated jointly with other undertakings, such as trade liberalization. This broader scope of the negotiations might open up possibilities for implicit or explicit side payments. To account for such possibilities, we assume here that the negotiating parties have access to unconstrained side payments. Specifically, Home and Foreign bargain over the vector $\mathbf{T}$ of general compensation schemes and a (possibly negative) side payment s from Home to Foreign to maximize the unconstrained Nash Product

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

where the parameter $\alpha \in (0,1)$ captures the host country bargaining power relative to that of the source country. Under unconstrained side payments, the parties agree on a compensation scheme $\mathbf{T}$ that maximizes their expected joint surplus $\omega(\mathbf{T})$, and negotiate the side payment s to determine the distribution of the surplus. By implication, the following proposition establishes circumstances under which the two contracting parties can do no better than to negotiate a carve-out scheme $\mathbf{T}^C$ if they have access to unconstrained side payments (see Appendix A.1 for a proof):

Proposition 1 *There exists a carve-out compensation scheme* $\mathbf{T}^C$ *that implements the jointly efficient outcome* $(\mathbf{k}^J, \theta^J)$ *if*

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

If condition (7) is violated, then $(\mathbf{k}^J, \theta^J)$ *cannot be implemented by any compensation scheme* $T^h(\mathbf{k}, \theta) \in [0, \Pi^h(\mathbf{k})]$ *for all* $h \in \mathcal{H}$.

The left-hand side of equation (7) measures the expected profit of firm h if all firms invest the jointly efficient portfolio $\mathbf{k}^J$ and firm h enjoys *complete investment protection*. The right-hand side measures the maximal expected profit firm h can obtain if all other firms invest efficiently, $\mathbf{k}_{-h}^J = (k_1^J, ..., k_{h-1}^J, ..., k_{h+1}^J, ..., k_H^J)$, and the agreement offers *no investment protection*. Condition (7) forms the basis of an incentive compatibility constraint on efficient investment under a carve-out policy. If this condition is satisfied for all investors, then there exists an equilibrium level of investment protection that we label $\theta^{JC} \equiv \Theta^C(\mathbf{k}^J)$, under which $\mathbf{k}^J$ is an equilibrium investment portfolio. Conversely, if (7) is violated, then the deviation profit is so large that at least one firm will deviate from k_h^J even if it does not receive any compensation at all subject to regulation. Observe that the problem of implementing $(\mathbf{k}^J, \theta^J)$ if (7) is violated extends beyond carve-out compensation schemes. No scheme that requires non-negative compensation and limits compensation to at most foregone operating profit can then implement $(\mathbf{k}^J, \theta^J)$. In that case, efficient implementation requires either negative $(T^h(\mathbf{k}, \theta) < 0)$ or punitive $(T^h(\mathbf{k}, \theta) > \Pi^h(\mathbf{k}))$ compensation payments in some states of the world. We will return to these issues below.

4.1.2 No side payments

Investment treaties are often negotiated separately from other state-to-state agreements. In this case, the possibilities for side payments are much more limited. To capture this feature, we consider the other polar case where the parties do not have access to any side payments. The contracting parties then negotiate $\mathbf{T}$ to maximize the constrained Nash Product

$$\mathcal{N}(\mathbf{T}) \equiv \hat{\mathcal{N}}(\mathbf{T}, 0) = [\tilde{V}(\mathbf{T}) - v^0]^\alpha [\tilde{\Pi}(\mathbf{T}) - \pi^0]^{1-\alpha}. \quad (8)$$

Joint surplus maximization does not follow trivially from bargaining in this case, because $\mathbf{T}$ is also required to distribute surplus across the contracting parties in accordance with their bargaining power α. Yet, the next proposition shows that the simple carve-out scheme $\mathbf{T}^C$ can nevertheless implement the jointly efficient outcome as well as distribute surplus according to the Nash Bargaining Solution under certain circumstances (the proof is in Appendix A.2):

Proposition 2 *An investment agreement with carve-out compensation* $\mathbf{T}^C$ *can implement the jointly* *efficient outcome* $(\mathbf{k}^J, \theta^J)$ *and maximize the Nash Product* $\mathcal{N}(\mathbf{T})$ *if and only if*

$$\max\{\Gamma(\mathbf{k}^J); F(\theta^J)\} \leq \frac{R(\mathbf{k}^J) + \pi^0}{\Pi(\mathbf{k}^J)} + (1-\alpha)\frac{\omega^J - \omega^0}{\Pi(\mathbf{k}^J)} \leq 1, \tag{9}$$

where

$$\Gamma(\mathbf{k}^J) \equiv \max_{h \in \mathcal{H}} \{\frac{\pi_h^d(\mathbf{k}_{-h}^J) + R^h(k_h^J)}{\Pi^h(\mathbf{k}^J)}\}. \tag{10}$$

By the construction of the variable $\Gamma(\mathbf{k}^J)$ defined in (10), condition (7) of Proposition 1 is met if and only if $\Gamma(\mathbf{k}^J) \leq 1$. However, this condition is not sufficient in the present context because the compensation scheme $\mathbf{T}^C$ must also distribute surplus across the two contracting parties if they have no side payments. Hence, the more restrictive condition (9). We return to this condition below where we illustrate how a carve-out compensation scheme can achieve the desired division of surplus. Conditions (7) and (9) are both violated if $\Gamma(\mathbf{k}^J) > 1$. Obviously, the two contracting parties cannot negotiate a jointly efficient outcome by way of a carve-out compensation scheme without side payments if they can not do so under unlimited side payments.

We examine the restrictiveness of condition (9) in light of the following two assumptions:

$$\textstyle\int_{\underline{\theta}}^{\theta^J} V(\mathbf{k}^J, \theta) dF(\theta) > v^0 \text{ and } \Pi(\mathbf{k}^J) - R(\mathbf{k}^J) > \pi^0. \tag{11}$$

Under the first assumption, the expected host country surplus is strictly higher under jointly efficient investment and regulation compared to the case of no agreement, gross of any expected compensation payments. If this assumption is violated, it is not possible to implement $(\mathbf{k}^J, \theta^J)$ and simultaneously satisfy the host country's participation constraint, even with a general compensation scheme. The second assumption ensures that the expected industry profit is strictly higher under jointly efficient investment and complete investment protection than with no agreement. If this assumption is violated, there exists no compensation scheme without punitive payments that can implement $(\mathbf{k}^J, \theta^J)$, and at the same time satisfy the source country's participation constraint.

Appendix A.2 verifies that (9) holds for bargaining strengths α in an interval $[\underline{\alpha}, \bar{\alpha}]$ if (11) holds and the industry consists of one single investor, $H = 1$. The Appendix also establishes plausible conditions under which (9) holds if the industry consists of $H \geq 2$ investors.

4.1.3 How carve-out compensation achieves efficiency and surplus distribution

We will first show that the incentives to regulate and invest are efficient for all investment protection above a certain 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 in full, and it regulates without compensation for $\theta > \theta^{JC}$, all of which is jointly efficient. 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* $\theta^{JC} \in [\theta^J, \bar{\theta}]$. 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 a carve-out scheme to implement the jointly efficient regulation subject to jointly efficient investment.

Investment incentives Investment protection $\Theta^C(\mathbf{k})$ under the carve-out scheme defined in (3) is a function of how much firms invest in the host country. Starting at the efficient investment portfolio $\mathbf{k}^J$, firm h contemplating a unilateral deviation $k_h \neq k_h^J$ must assess how investment protection will change with a change in investment. The carve-out compensation scheme disciplines investment precisely through the related change in investment protection. Consider the extreme case in which $\mathbf{T}^C$ offers no investment protection if one or more firms deviate from the efficient portfolio: $\Theta^C(\mathbf{k}) = \Theta(\mathbf{k})$ for all $\mathbf{k} \neq \mathbf{k}^J$. This is akin to a grim trigger strategy.[19] Firm h's maximal deviation profit is then given by $\pi^d(\mathbf{k}_{-h}^J)$ defined in (7). To ensure incentive compatibility, $\mathbf{T}^C$ must offer sufficient investment protection θ^{JC} in equilibrium to render a deviation unprofitable:

$$F(\theta^{JC})\Pi^h(\mathbf{k}^J) - R^h(k_h^J) \geq \pi_h^d(\mathbf{k}_{-h}^J).$$

This condition must hold for all H investors, which yields a lower bound

$$F(\theta^{JC}) \geq \max_{h \in \mathcal{H}} \{ \frac{\pi^d(\mathbf{k}_{-h}^J) + R^h(k_h^J)}{\Pi^h(\mathbf{k}^J)} \} = \Gamma(\mathbf{k}^J)$$

on the equilibrium investment protection θ^{JC}. This mechanism, while extreme, illustrates the general point that *anticipated changes in investment protection* $\Theta^C(\mathbf{k})$ can incite strategic investors to behave efficiently.

Surplus division The incentive compatibility constraints—$\theta^{JC} \geq \theta^J$ for regulation and $\theta^{JC} \geq F^{-1}(\Gamma(\mathbf{k}^J))$ for investment—establish two *lower bounds* on the equilibrium investment protection under carve-out compensation, so that any value of θ^{JC} that exceeds the larger of those two bounds will implement the jointly efficient outcome. Such investment protection exists if condition (7) is satisfied because then $\Gamma(\mathbf{k}^J) \leq 1$ and therefore $F^{-1}(\Gamma(\mathbf{k}^J)) \leq \bar{\theta}$. If the contracting parties have access to unconstrained side payments, they can divide surplus by way of those side payments. If they cannot implement direct side payments, they must resort to other measures. Varying the investment protection θ^{JC} within the feasible range $[\max\{F^{-1}(\Gamma(\mathbf{k}^J)); \theta^J\}, \bar{\theta}]$ offers such a possibility.

Under a carve-out compensation scheme $\mathbf{T}^C$ that implements $(\mathbf{k}^J, \theta^J)$, the host country equi-

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

librium expected surplus is

$$\tilde{V}(\mathbf{T}^C) \equiv \int_{\underline{\theta}}^{\theta^J} V(\mathbf{k}^J,\theta)dF(\theta) - [F(\theta^{JC}) - F(\theta^J)]\Pi(\mathbf{k}^J), \tag{12}$$

whereas the equilibrium expected industry profit equals

$$\tilde{\Pi}(\mathbf{T}^C) = F(\theta^{JC})\Pi(\mathbf{k}^J) - R(\mathbf{k}^J). \tag{13}$$

By substituting these expressions into (8), we obtain the Nash Product

$$\mathcal{N}(\mathbf{T}^C) = [\int_{\underline{\theta}}^{\theta^J} V(\mathbf{k}^J,\theta)dF(\theta) - (F(\theta^{JC}) - F(\theta^J))\Pi(\mathbf{k}^J) - v^0]^\alpha [F(\theta^{JC})\Pi(\mathbf{k}^J) - R(\mathbf{k}^J) - \pi^0]^{1-\alpha}$$

under a carve-out compensation scheme $\mathbf{T}^C$ that implements the jointly efficient outcome $(\mathbf{k}^J, \theta^J)$. In particular, a higher θ^{JC} transfers expected surplus from Home to Foreign by way of higher expected compensation payments. Investment protection effectively is an indirect side payment. By maximizing $\mathcal{N}(\mathbf{T}^C)$ over θ^{JC}, we solve for the level of investment protection

$$F(\theta^{JC}) = \frac{R(\mathbf{k}^J) + \pi^0}{\Pi(\mathbf{k}^J)} + (1-\alpha)\frac{\omega^J - \omega^0}{\Pi(\mathbf{k}^J)} \tag{14}$$

that maximizes the Nash Product $\mathcal{N}(\mathbf{T}^C)$. 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. The investment protection characterized in (14) is contained in $[\max\{F^{-1}(\Gamma(\mathbf{k}^J)); \theta^J\}, \bar{\theta}]$ if and only if condition (9) of Proposition 2 holds.

4.1.4 Implications for the interpretation of core features of agreements

Propositions 1 and 2 show 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. These results thus give 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. Propositions 1 and 2 show 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 agreements derived above include 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. Direct support of foreign investment occurs in practice through general subsidy schemes, and in government contracts with specific firms, but is not an integral 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 overregulation, 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 outcome, and also serve as implicit side payments in the negotiation of the investment agreement.*

There is no need for ex ante subsidy commitments when carve-out compensation implements the efficient outcome. In other circumstances, investment subsidies can potentially improve efficiency; see Section 4.3.

4.2 When carve-out compensation cannot implement the efficient outcome

Propositions 1 and 2 identify necessary and sufficient conditions for when negotiations over carve-out compensation schemes will implement the jointly efficient outcome. By implication they also identify circumstances under which carve-out compensation mechanisms cannot be expected to implement ($\mathbf{k}^J$,θ^J). 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 condition (7) is violated. 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. In this case, condition (9) is violated 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.3):

Proposition 3 *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 surplus (*$\tilde{V}(\mathbf{T}^C) \geq \tilde{V}(\mathbf{T}^N)$*).*

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.3 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 3 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 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 apparent from Proposition 1 is an agreement that allows 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.[20] We do 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.[21]

Firm-specific compensation A third possibility could be to give firms *different fractions* of their foregone operating profits. Firm-specific θ_h^{JC} 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 surplus of entering into an agreement implementing $(\mathbf{k}^J, \theta^J)$.

Compensation based on other variables than foregone operating profits In our model,

[20] 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 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.

[21] Aisbett, Karp and McAusland (2010a) show how efficiency can be implemented using a compensation scheme that 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, however: $\beta(\mathbf{k}^J) > 1$. They also show how compensation based on a linear combination of operating profits $\Pi^h(\mathbf{k})$ and investment costs $R^h(\mathbf{k})$ can achieve efficiency. This mechanism is efficient because it has two instruments that can be used to incite investment and regulation, but it still involves excessive compensation.

pay-off relevant variables that govern investment and regulation include investment costs $R(\mathbf{k})$ and host country surplus $V(\mathbf{k},\theta)$. 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.4):

Proposition 4 *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 surplus (*$\tilde{V}(\hat{\mathbf{T}}) \geq \tilde{V}(\mathbf{T})$*).*

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 increase in regulatory efficiency. The intuition is analogous to that 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.[22] 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 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 4 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

[22] 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.

$-V(\mathbf{k}^J, \theta)$ 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 1 *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.*

4.3 Extensions

This section describes six modifications to our baseline model. Formal statements and their verifications are in Appendix A.5.

Non-contingent investment protection In the carve-out compensation scheme defined in (3), investment protection is a function of investment. Propositions 1 and 2 rely on sophisticated behavior in the sense that investment incentives are shaped by beliefs about how changes in investment affect investment protection $\Theta^C(\mathbf{k})$. Appendix A.5.1 considers an agreement in which investment protection is *independent* of investment, and establishes sufficient conditions for when such an agreement can implement the jointly efficient outcome. If the industry consists of multiple asymmetric firms, efficient implementation requires firm-specific investment protection. But implementation of the Nash Bargaining Solution generally requires that contracting parties have access to side payments. If not, the negotiated agreement involves a trade-off between efficiency and surplus distribution.

Two-way investment flows In the above analysis, all investments flow from Foreign to Home. But countries increasingly enter into agreements to stimulate both their in- and outward investments. Similarly, existing agreements that have effectively applied to investment from developed to developing countries only, are increasingly applying also to investments in the opposite direction due to the economic development of the developing country partners. Appendix A.5.2 therefore considers agreements with *two-way investments*. A negotiation here concerns a pair of compensation mechanisms, one for each country. A main finding is that carve-out compensation can implement the jointly efficient outcome through Nash Bargaining in a broader set of circumstances than with separate agreements for the two investment directions. The reason is that investment protection levels in the two countries are perfect complements regarding surplus distribution in an agreement covering two-way investment. This also implies that symmetric investment protection can implement the Nash Bargaining Solution even in circumstances where countries are asymmetric.

Multiple industries A defining characteristic of actual investment agreements is their economy-wide scope. To highlight a bargaining-related benefit of this feature, Appendix A.5.3 examines a setting in which *multiple industries* are exposed to separate shocks θ_i. Carve-out compensation can then implement the jointly efficient outcome with Nash Bargaining in a broader set of circumstances than with industry-specific agreements since investment protection levels in the different industries are perfect substitutes regarding surplus distribution in an economy-wide agreement.

Partial regulation Our baseline model assumes that the host country has a binary choice between allowing and disallowing production. This is intended to capture the standard case law interpretation that a policy measure must wipe out almost all profit to constitute regulatory expropriation.[23] Appendix A.5.4 considers a model of *partial regulation,* where the extent of policy intervention is a continuous choice, with no regulation and complete shut-down as polar extremes. It is not obvious how to define a compensation scheme with partial regulation. But the Appendix shows that a scheme that either fully undoes all negative consequences of regulation (relative to the efficient benchmark) or offers no compensation, can implement the jointly efficient outcome under partial regulation under similar conditions to when this can be done with a binary choice of regulation and carve-out compensation. However, with partial regulation it is not possible to implement the Nash Bargaining Solution with the carve-out scheme, contrary to what is possible in the setting underlying Proposition 2.

Investment subsidies Our analysis has excluded investment subsidies or tax breaks since they are hardly ever included in investment agreements. Appendix A.5.5 examines consequences of including *investment subsidies,* and shows that agreements based on carve-out compensation then are Pareto optimal. Investment subsidies can be used both to stimulate investment and distribute surplus, which leaves efficient regulation as the only objective of the compensation scheme. A simple carve-out scheme in which investment protection is the same for all firms and independent of investment is then sufficient to align host country incentives.

Imperfect enforcement The fundamental task of an arbitration court is to decide whether a challenged measure constitute expropriation, and if so, how much compensation investors should receive. Under a carve-out mechanism and efficient investment, this task reduces to assessing whether the regulatory shock was mild in the sense of $\theta \leq \theta^{JC}$, in which case each investor h should

[23] More relevant for cases where investors are deprived of less than the full value of their investment is the fair-and-equitable treatment provision included in many agreements. This amorphous provision plays a central role in case law, but has not been subject to any formal economic analysis, to the best of our knowledge.

be paid its foregone operating profit $\Pi^h(\mathbf{k}^J)$. Solving this task is straightforward by the assumptions that θ and $\Pi^h(\mathbf{k}^J)$ are verifiable.

In Appendix A.5.6, the arbitration court instead correctly identifies regulatory expropriation with a probability $Q(\theta^{JC} - \theta)$ that is higher, the smaller is the regulatory shock θ, and the more extensive is investment protection (θ^{JC}). Perfect enforcement in the sense of perfect verifiability of all regulatory shock is not fundamental for implementation of the jointly efficient outcome. Instead, $Q(\theta^{JC} - \theta^J) = 1$ is necessary and sufficient to prevent overregulation—the arbitration court only needs to be able to identify cases where regulations occurs for shocks $\theta \leq \theta^J$.

Investor losses can also be hard to verify, although arbitration courts often rely on external accounting expertise. We show in Appendix A.5.6 that compensation mechanisms cannot be used to elicit any information about foregone operating profit. Compensation schemes must build on some external benchmark, such as expected operating profit if audits cannot verify investor losses.

5 Policy issues

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.

5.1 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* occurs 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 4, 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 efficient, for shocks $\theta > \Theta^J(\mathbf{k})$. Proposition 4 directly implies:

Corollary 2 *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 efficient to regulate. We explained the intuition for this result in the discussion following Proposition 4. 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 efficient.[24] 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.

5.2 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 their domestic legal and regulatory frameworks. Additionally, the agreements are meant to stimulate investment flows in *both* directions. We will refer to these as *North-North agreements.* The 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.

5.2.1 The different rationales of the agreements

To see the importance of South's lack of commitment ability for the scope of a North-South agreement, suppose South can commit to investment protection even without an agreement, but invest-

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

ment 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 $\tilde{V}(\mathbf{T})$. 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 joint 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 $\omega(\mathbf{T}^C) > \omega^U$. 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 North-North countries, the joint expected surplus generated in Home is $\tilde{V}(\mathbf{T}^U) + \tilde{\Pi}(\mathbf{T}^U)$ absent an investment agreement. Since Home ignores the positive effect of its investment protection on Foreign expected profit $\tilde{\Pi}(\mathbf{T}^U)$, θ^U is too small from a joint surplus perspective.[25] The same is true for the equilibrium investment protection θ^{*U} in Foreign under the carve-out compensation scheme $\mathbf{T}^{*U}$ that maximizes the expected surplus $\tilde{V}^*(\mathbf{T}^*)$ in Foreign. Absent an 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 correct 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. In sum:

Observation 2 *Concerning the scope for an investment agreement:*
(1) A host country that can unilaterally implement any compensation mechanism to protect invest-

[25] This will not hold if Home can maximize joint surplus and extract it all through a side payment. In that case, the only motive for an agreement with two-way investment flows would be to solve domestic commitment problems.

ment 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 domestic investment protection.

A substantial 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. The symmetric nature of the contracting parties suggests that these agreements are meant to promote two-way investment. In these cases, there is a double benefit to an agreement: improved enforcement through international arbitration tribunals and internalization of international policy externalities. The Nash Product characterized in (A.10) is specified under the assumption of no investment protection in either country absent an agreement. If Home instead has unilateral commitment possibilities, then the status quo points in (A.10) change from v^0 and π^0 to v^U and π^U. Compared to the benchmark of no investment protection, the higher national surplus in Home absent an agreement, $v^U > v^0$, improves Home's bargaining position, which tends to reduce the negotiated investment protection on inward investment and/or increase the negotiated investment protection on outward investment. However, the higher expected profit on outward investment absent an agreement, $\pi^U > \pi^0$, has the opposite effect on investment protection through an improvement in Foreign's bargaining position. The total effect on investment protection of improving domestic enforcement possibilities in Home is ambiguous. By implication, South-South agreements may feature larger or smaller carve-outs than North-North agreements.

5.2.2 The different distributional impacts of the agreements

In a North-South agreement, both parties prefer 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 surplus $\tilde{V}(\mathbf{T}^C)$ that is generated domestically by inward investment from Foreign. The second source is the expected industry profit $\tilde{\Pi}^*(\mathbf{T}^{*C})$ of Home investment in Foreign.[26] Home and Foreign have a common interest to let investment protection be at least at the unilaterally optimal levels (θ^U, θ^{*U}), but both levels 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 $\pi^{*U} = \tilde{\Pi}^*(\mathbf{T}^{*U})$

[26] See Appendix A.5.2 for a more detailed analysis of an investment agreement with two-way investment flows.

is the expected industry profit from Home's investment in Foreign if there is no agreement, and Foreign commits to carve-out compensation $\mathbf{T}^{*U}$. By implication:

$$\tilde{\Pi}^*(\mathbf{T}^{*C}) - \pi^{*U} > v^U - \tilde{V}(\mathbf{T}^C) > 0. \tag{15}$$

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

Proposition 5 *Concerning the distribution of surplus in investment agreements:*
(1) North-South and North-North agreements entail more investment protection than the levels that maximize the expected domestic surpluses.
(2) A North-South agreement benefits investors from North and increases expected domestic surplus in South.
(3) A North-North agreement benefits foreign investors in both countries, but reduces the expected domestic surplus in each country.

We believe that these results might shed some 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. The EU and the US are in all likelihood capable of providing any level of protection of foreign investment that they prefer through their existing legal systems. Since our analysis does not distinguish between different factor owners, and does not describe the general equilibrium implications of the formation of investment agreements, we cannot precisely identify the distributional impacts. But our framework suggests that the direct effect of the additional investment protection offered by these agreements would mainly benefit foreign investors and harm the rest of society. To the extent that public opinion is mainly determined by these direct effects, the agreements will *always appear too protective of foreign investor interests, from the point of view of the rest of society.*[27]

6 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

[27] 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 (15).

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.

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). The 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 left aside. 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 other 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 interpretation. 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 (2018). "North-South FDI and Bilateral Investment Treaties." *World Economy* 41(1), 2-28.

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

Horn, Henrik and Thomas Tangerås (2017). "Economics and Politics of International Investment Agreements." Centre for Economic Policy Research Discussion Paper No. 11 879 (revised version with the title "Economics of International Investment Agreements).

Horn, Henrik and Thomas Tangerås (2020). "Investor-State vs. State-State Dispute Settlement." Centre for Economic Policy Research, Discussion Paper no. 14480.

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." *Journal of International Economics* 121, 103254.

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.

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 (2020): "Disputes in International Investment and Trade." NBER Working Paper 27012.

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

Schjelderup, Guttorm and Frank Stähler (2020). "Investor State Dispute Settlement and Multinational Firm Behavior." Department of Business and Management Science Discussion Paper 9, Norwegian School of Economics.

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.

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.

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

Proof: $V(\mathbf{k}, \theta)$ 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* $\Theta^C(\mathbf{k}) \geq \Theta(\mathbf{k})$.[28] *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 $\theta > \hat{\Theta}(\mathbf{k})$, it is strictly optimal to regulate:

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

It is also strictly optimal to regulate if $\hat{\Theta}(\mathbf{k}) = \Theta^J(\mathbf{k})$ and $\theta > \hat{\Theta}(\mathbf{k})$:

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

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

We first show that the negotiated agreement maximizes $\omega(\mathbf{T})$. Holding $\mathbf{T}$ fixed, $\hat{\mathcal{N}}(\mathbf{T}, s)$ is strictly quasi-concave in s and reaches its global optimum at

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

Moreover,

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

implies that the agreement maximizes $\omega(\mathbf{T})$.

To see that the contracting parties will choose an agreement that implements $(\mathbf{k}^J, \theta^J)$ if such an agreement is feasible, note that for any $\mathbf{T}$,

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

because regulation is ex post efficient if $\Theta^J(\hat{\mathbf{k}})$ defines the threshold for regulation. Since $\mathbf{k}^J$ maxi-

[28] 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.

mizes $\Omega(\mathbf{k})$, $\omega^J = \Omega(\mathbf{k}^J) \geq \Omega(\hat{\mathbf{k}}) \geq \omega(\mathbf{T})$ for all $\mathbf{T}$.

Sufficiency of condition (7). This condition is equivalent to:

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

By implication, $F^{-1}(\Gamma(\mathbf{k}^J)) \leq \bar{\theta}$. Consider a $\mathbf{T}^C$ where investment protection $\Theta^C(\mathbf{k})$ satisfies

$$\Theta^C(\mathbf{k}) = \begin{cases} \theta^{JC} \in [\max\{F^{-1}(\Gamma(\mathbf{k}^J)); \theta^J\}, \bar{\theta}] & \text{if } \mathbf{k} = \mathbf{k}^J \\ \Theta(\mathbf{k}) & \text{if } \mathbf{k} \neq \mathbf{k}^J \end{cases} . \tag{A.3}$$

Under $\mathbf{T}^C$, 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 $\mathbf{k}_{-h}^J = (k_1^J, ...k_{h-1}^J, k_{h+1}^J, ..., k_H^J)$ by $\Theta^C(\mathbf{k}^J) = \theta^{JC} \geq \theta^J = \Theta^J(\mathbf{k}^J) > \Theta(\mathbf{k}^J)$ and Lemma 2. If h unilaterally deviates to $k_h \neq k_h^J$, then its expected profit becomes instead

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

As unilateral deviations are unprofitable for all firms, it follows that $\mathbf{T}^C$ with investment protection (A.3) implements $\mathbf{k}^J$. $\mathbf{T}^C$ also implements efficient regulation in equilibrium by $\Theta^C(\mathbf{k}^J) = \theta^{JC} \geq \theta^J = \Theta^J(\mathbf{k}^J)$ and Lemma 2.

Necessity of condition (7). We prove necessity by showing that condition (7) 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 $\pi_h^d(\mathbf{k}_{-h}^J)$ is

$$\begin{aligned} & \Pi^h(k_h^d, \mathbf{k}_{-h}^J) \textstyle\int_{M(k_h^d, \mathbf{k}_{-h}^J)} dF(\theta) + \int_{M^r(k_h^d, \mathbf{k}_{-h}^J)} T^h(k_h^d, \mathbf{k}_{-h}^J, \theta) dF(\theta) - R^h(k_h^d) \\ = \quad & \Pi^h(k_h^d, \mathbf{k}_{-h}^J)[\textstyle\int_{M(k_h^d, \mathbf{k}_{-h}^J)} dF(\theta) - F(\Theta(k_h^d, \mathbf{k}_{-h}^J))] + \int_{M^r(k_h^d, \mathbf{k}_{-h}^J)} T^h(k_h^d, \mathbf{k}_{-h}^J, \theta) dF(\theta) + \pi_h^d(\mathbf{k}_{-h}^J) \\ \geq \quad & \pi_h^d(\mathbf{k}_{-h}^J). \end{aligned}$$

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 $\pi_h^d(\mathbf{k}_{-h}^J)$ under a deviation from k_h^J under any compensation mechanism with non-negative compensation.

A deviation is strictly profitable for at least one firm if condition (7) is violated.■

A.2 Proof of Proposition 2

Sufficiency of condition (9). The equilibrium investment protection $\theta^{JC} = \Theta^C(\mathbf{k}^J)$ defined in (14) satisfies $\theta^{JC} \in [\max\{F^{-1}(\Gamma(\mathbf{k}^J)); \theta^J\}, \bar{\theta}]$. Hence, a carve-out policy $\mathbf{T}^C$ with investment protection $\Theta^C(\mathbf{k})$ defined in (A.3) implements $(\mathbf{k}^J, \theta^J)$; see the proof of Proposition 1. Substituting this compensation scheme into the Nash Product defined in (8) yields for all $\mathbf{T}$:

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

Necessity of condition (9). The threshold θ^{JC} defined in (14) uniquely characterizes the equilibrium investment protection that maximizes $\mathcal{N}(\mathbf{T}^C)$ under implementation of $(\mathbf{k}^J, \theta^J)$. If

$$\Gamma(\mathbf{k}^J) > \frac{R(\mathbf{k}^J) + \pi^0}{\Pi(\mathbf{k}^J)} + (1-\alpha)\frac{\omega^J - \omega^0}{\Pi(\mathbf{k}^J)},$$

then

$$\pi_h^d(\mathbf{k}_{-h}^J) > F(\theta^{JC})\Pi^h(\mathbf{k}^J) - R^h(k_h^J)$$

for at least one firm, in which case k_h^J is not incentive compatible. If

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

then $\theta^{JC} < \theta^J$, in which case the host country will overregulate for all $\theta \in (\theta^{JC}, \theta^J)$ under efficient investment $\mathbf{k}^J$. If

$$\frac{R(\mathbf{k}^J) + \pi^0}{\Pi(\mathbf{k}^J)} + (1-\alpha)\frac{\omega^J - \omega^0}{\Pi(\mathbf{k}^J)} > 1,$$

then $\theta^{JC} > \bar{\theta}$, which is infeasible.

Robustness. Assume that (11) holds. In this case $\frac{R(\mathbf{k}^J)+\pi^0}{\Pi(\mathbf{k}^J)} < 1$, so a necessary and sufficient condition for (9) to hold for an interval $\alpha \in [\underline{\alpha}, \bar{\alpha}]$, $0 \leq \underline{\alpha} < \bar{\alpha} \leq 1$, is

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

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

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

If $\Gamma(\mathbf{k}^J) > F(\theta^J)$, then we can rewrite the above condition as

$$\omega^J - \omega^0 > \Gamma(\mathbf{k}^J)\Pi(\mathbf{k}^J) - R(\mathbf{k}^J) - \pi^0. \tag{A.4}$$

If $H = 1$, then the right-hand side of (A.4) is zero because then $\pi_h^d(\mathbf{k}_{-h}^J) = \pi^0$, and we are done. To evaluate (A.4) for $H \geq 2$, we add more structure, by assuming that all firms are symmetric. (A.4) then 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 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 then $\pi^d < \pi^0$. These conditions are sufficient, but not necessary. By continuity, (A.4) 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.3 Proof of Proposition 3

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 surplus. As the marginal benefit 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 (6). 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 $V(\hat{\mathbf{k}}, \hat{\theta})$ inside each of the two integrals and rewrite:

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

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.4 Proof of Proposition 4

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 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

$$F(\hat{\Theta}(\mathbf{k})) \equiv \min\{\textstyle\int_{M(\mathbf{k})} dF(\theta); F(\Theta^J(\mathbf{k}))\} \tag{A.5}$$

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

$$\hat{T}^h(\mathbf{k}, \theta) = \begin{cases} \Pi^h(\mathbf{k}) & \theta \in A(\mathbf{k}) \cup A^r(\mathbf{k}) = [\underline{\theta}, \hat{\Theta}(\mathbf{k})] \\ \tilde{T}^h(\mathbf{k}, \theta) & \theta \in B(\mathbf{k}) \\ T^h(\mathbf{k}, \theta) & \theta \in B^r(\mathbf{k}) \end{cases}. \tag{A.6}$$

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

$$\tilde{T}^h(\mathbf{k},\theta) \equiv \frac{1}{\int_{B(\mathbf{k})} dF(\tilde{\theta})}[\int_{A^r(\mathbf{k})} T^h(\mathbf{k},\tilde{\theta})dF(\tilde{\theta}) + \max\{\int_{M(\mathbf{k})} dF(\tilde{\theta}) - F(\Theta^J(\mathbf{k})); 0\}\Pi^h(\mathbf{k})]. \quad \text{(A.7)}$$

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.5). 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 $\hat{\Theta}(\mathbf{k})$:

$$\textstyle\int_{B(\mathbf{k})} dF(\tilde{\theta}) \equiv \int_{A^r(\mathbf{k})} dF(\tilde{\theta}) + \max\{\int_{M(\mathbf{k})} dF(\tilde{\theta}) - F(\Theta^J(\mathbf{k})); 0\}. \quad \text{(A.8)}$$

Use $\tilde{T}^h(\mathbf{k},\theta)$ defined in (A.7), and (A.8) 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.8) 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.5) and the compensation rules (A.6)-(A.7), the expected investment profit of every firm $h \in \mathcal{H}$ is the same under both compensation mechanisms for all $\mathbf{k}$:

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

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 effect of compensation payments on the surplus 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 (6). 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 $V(\hat{\mathbf{k}},\hat{\theta})$ inside the two integrals and rewriting yields

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

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.6) and (A.7), 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 $\hat{\mathbf{T}}$ is therefore a convex combination of operating profit and compensation under the original agreement $\mathbf{T}$, where the weights depend on $(\mathbf{k},\theta)$, but are the same for all firms. This property implies that the modified scheme $\hat{\mathbf{T}}$ inherits a number of characteristics from the initial scheme **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 4 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.5 Verification of statements in Section 4.3

A.5.1 Non-contingent investment protection

Let an industry consist of a single foreign firm, $H = 1$, with operating profit $\Pi(k) \geq 0$ and investment cost $R(k) \geq 0$. Assume that both expressions are strictly increasing and well-behaved ($\Pi_{kk}(k) < 0$, $R_{kk}(k) \geq 0$). Let k^0 be the equilibrium investment absent any agreement and $\theta^0 = \Theta(k^0)$ the corresponding threshold for regulation. Denote by (k^J, θ^J) the outcome that maximizes the expected joint surplus of the host country and the foreign investor, where $\theta^J = \Theta^J(k^J)$.

Consider an agreement in which investment protection is independent of investment k, under which the investor expects to retain its full operating profit for $\theta \leq \hat{\theta}$, and be regulated without compensation for $\theta > \hat{\theta}$. The profit-maximizing investment is

$$K(\hat{\theta}) \equiv \arg\max_{k \geq 0} \{F(\hat{\theta})\Pi(k) - R(k)\},$$

with $K_\theta(\hat{\theta}) > 0$ for $K(\hat{\theta}) > 0$.

Proposition A.1 *Assume that a single foreign firm invests in an industry ($H = 1$). Then there exists a θ^{JC} such that the carve-out compensation function*

$$T^C(k, \theta, \theta^{JC}) \equiv \begin{cases} \Pi(k) & \text{if } \theta \leq \theta^{JC} \\ 0 & \text{if } \theta > \theta^{JC} \end{cases} \tag{A.9}$$

implements the jointly efficient outcome (k^J, θ^J) if $k^J \in [K(\theta^J), K(\bar{\theta})]$ and $\theta^J > \theta^0$.

Proof: Define the subset $\kappa(\hat{\theta}) \equiv \{k \geq 0 : \Theta(k) \leq \hat{\theta}\}$ and its complement $\kappa^c(\hat{\theta}) \equiv \{k \geq 0 : \Theta(k) > \hat{\theta}\}$. By way of these definitions, the firm has expected investment profit $F(\hat{\theta})\Pi(k) - R(k)$ for all $k \in \kappa(\hat{\theta})$ and $F(\Theta(k))\Pi(k) - R(k)$ for all $k \in \kappa^c(\hat{\theta})$ under the carve-out compensation scheme $T^C(k, \theta, \hat{\theta})$, $\hat{\theta} \in [\underline{\theta}, \bar{\theta}]$. If $k^J \in [K(\theta^J), K(\bar{\theta})]$, then $K(\theta^{JC}) = k^J$ for some $\theta^{JC} \in [\theta^J, \bar{\theta}]$ by the mean-value theorem.

Consider the profit-maximizing investment under the carve-out compensation scheme $T^C(k, \theta, \theta^{JC})$. Observe that $k^J \in \kappa(\theta^{JC})$ by $\Theta(k^J) < \theta^J \leq \theta^{JC}$. Since $k^J = K(\theta^{JC})$ maximizes $F(\theta^{JC})\Pi(k) - R(k)$ over $k \geq 0$, k^J also constitutes a profit-maximizing investment in the subset $\kappa(\theta^{JC})$. Consider next a deviation to $k \in \kappa^c(\theta^{JC})$. If we let $\pi^0 = \max_{k\geq 0}\{F(\Theta(k))\Pi(k) - R(k)\}$, then $\pi^0 \geq F(\Theta(k))\Pi(k) - R(k)$ for all $k \in \kappa^c(\theta^{JC})$. Moreover, $F(\theta^{JC})\Pi(k^J) - R(k^J) \geq F(\theta^{JC})\Pi(k^0) - R(k^0) > F(\theta^0)\Pi(k^0) - R(k^0) = \pi^0$, where the weak inequality follows from optimality of k^J and the strict inequality from $\theta^{JC} \geq \theta^J > \theta^0$. By combining these inequalities, we obtain $F(\theta^{JC})\Pi(k^J) - R(k^J) > F(\Theta(k))\Pi(k) - R(k)$ for all $k \in \kappa^c(\hat{\theta})$, which establishes k^J as the profit-maximizing investment under $T^C(k, \theta, \theta^{JC})$.

Consider the ex post efficient regulation under the carve-out compensation scheme $T^C(k, \theta, \theta^{JC})$. Assume that the firm has invested k^J. Since $\Theta(k^J) < \theta^J \leq \theta^{JC}$, it follows that the host country optimally allows production for all $\theta \leq \theta^J$ and regulates for all $\theta > \theta^J$.

If the firm is instead non-strategic, we can simply set $\kappa^c(\theta^{JC}) = \emptyset$ in the above proof.■

Investment protection is independent of investment in (A.9). A necessary condition for this scheme to implement the jointly efficient outcome is that underinvestment will occur if investment protection is too low, in the sense that $K(\theta^J) \leq k^J$.[29] Otherwise, all protection levels $\theta^{JC} \geq \theta^J$ will lead to overinvestment. A second necessary condition is that full investment protection will lead to overinvestment, $K(\bar{\theta}) \geq k^J$.[30] Otherwise, even full investment protection cannot implement k^J. If the jointly efficient solution in addition features less regulation than the benchmark without an agreement, $\theta^J > \theta^0$, then a carve-out compensation scheme in which investment protection is independent of the level of investment is sufficient to implement (k^J, θ^J).

Proposition A.1 extends to an industry with multiple foreign investors, $H \geq 2$, if these investors are symmetric. Otherwise, efficient implementation would require different levels of investment protection θ_h^{JC} for each firm h. Implementation of (k^J, θ^J) exploits all the degrees of freedom of $T^C(k, \theta, \theta^{JC})$. Implementation of the Nash Bargaining Solution therefore generally requires that the contracting parties have access to side payments. If not, the negotiated agreement will involve a trade-off between efficiency and surplus distribution. Note that the proof of Proposition A.1 also holds if the investor treats investment protection as exogenous to the own investment.

A.5.2 Two-way investment flows

Assume that Home and Foreign serve as both hosts and sources of foreign investment, but the profits from investment abroad are unrelated to activities in the domestic economy in each country. Negotiation over an agreement now concerns a pair of general compensation mechanisms $(\mathbf{T}, \mathbf{T}^*)$,

[29] A necessary and sufficient conditions for when $K(\theta^J) \leq k^J$ is that the host country expected surplus is increasing in investment when evaluated at (k^J, θ^J): $\int_{\underline{\theta}}^{\theta^J} V_k(k^J, \theta) dF(\theta) \geq 0$.

[30] Complete investment protection leads to overinvestment, $K(\bar{\theta}) \geq k^J$, if and only if $(1 - F(\theta^J))\Pi_k(k^J) \geq \int_{\underline{\theta}}^{\theta^J} V_k(k^J, \theta) dF(\theta)$. If the right-hand side is positive, then this condition is satisfied if the marginal operating profit evaluated at k^J is sufficiently large or if the jointly efficient threshold for regulation θ^J is sufficiently small.

where variables pertaining to Foreign are indicated by an asterisk (*). The outcome of the negotiations is given by the Nash Bargaining Solution, and the parties do not have access to side payments. The negotiated agreement thus maximizes

$$\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*) \equiv [\tilde{V}(\mathbf{T}) + \tilde{\Pi}^*(\mathbf{T}^*) - v^0 - \pi^{*0}]^\alpha [\tilde{V}^*(\mathbf{T}^*) + \tilde{\Pi}(\mathbf{T}) - v^{*0} - \pi^0]^{1-\alpha}. \tag{A.10}$$

The countries can alternatively negotiate two separate agreements over one-way investment flows. But since they 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 incentive compatibility condition (7), still applies in both countries under two-way investment. Even so, the agreement covering two-way investment flows is typically *more* efficient:

Proposition A.2 *Concerning investment agreements with two-way investment flows:*
(1) A carve-out scheme $(\mathbf{T}^C, \mathbf{T}^{*C})$ *implements the jointly efficient outcome* $(\mathbf{k}^J, \theta^J, \mathbf{k}^{*J}, \theta^{*J})$ *and maximizes* $\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*)$ *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.*

Proof: Assume that (7) holds in both countries (with appropriate asterisk notation for Foreign). Then all investment protection levels $(\theta^{JC}, \theta^{*JC})$ that satisfy

$$F(\theta^{JC}) \in [\max\{\Gamma(\mathbf{k}^J); F(\theta^J)\}, 1] \text{ and } F^*(\theta^{*JC}) \in [\max\{\Gamma^*(\mathbf{k}^{*J}); F^*(\theta^{*J})\}, 1]$$

implement $(\mathbf{k}^J, \theta^J)$ and $(\mathbf{k}^{*J}, \theta^{*J})$ both under joint and separate negotiations. 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

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

Inserting this expression into the Nash product $\mathcal{N}^B(\mathbf{T}, \mathbf{T}^*)$ above 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 (14), and θ^{*JC} in Foreign, given by

$$F^*(\theta^{*JC}) = \frac{R^*(\mathbf{k}^{*J})+\pi^{*0}}{\Pi^*(\mathbf{k}^{*J})} + \alpha\frac{\omega^{*J}-\omega^{*0}}{\Pi^*(\mathbf{k}^{*J})}, \tag{A.11}$$

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 $\Gamma^*(\mathbf{k}^{*J}) \leq F^*(\theta^{*J})$ and $\theta^{*JC} = \theta^{*J} - \varepsilon$ in (A.11). 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 (14) satisfies $\max\{\Gamma(\mathbf{k}^J);F(\theta^J)\} < F(\theta^{JC}) < 1$. 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 show that $\theta^{JC} = \theta^{*JC}$ sometimes is feasible also under asymmetries, start with symmetric countries and assume that $\max\{\Gamma(\mathbf{k}^J);F(\theta^J)\} < F(\theta^{JC}) < 1$. Introduce a small asymmetry, so that $F(\theta^{JC}) \neq F^*(\theta^{*JC})$ if defined by (14) and (A.11). 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}.■

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. The level of protection does not matter to the two countries as long as it does 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. This implies that countries can achieve the desired distribution of surplus in a *broader* set of circumstances than was feasible under one-way investment flows.

Turning to Part (2) of Proposition A.2, the jointly efficient 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 A.2 shows that such *contractual symmetry does not necessarily imply contractual inefficiency even if countries are asymmetric*. Formally, 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.

A.5.3 Multiple industries

Extend the baseline model of one-directional investment to allow for multiple of industries $i \in \{1, ..., I\}$. Each industry 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 surplus 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. If the carve-out compensation scheme implements $(\mathbf{k}_i^J, \theta_i^J)$ for all industries, then the host country expected surplus 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 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 the desired distribution of surplus:

Proposition A.3 *A carve-out compensation scheme implements the jointly efficient 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.*

A.5.4 Partial regulation

Assume that investment is one-directional and that there is a single potential foreign investor. Host country surplus now equals $V(k, \theta, y)$, and the investor operating profit is $\Pi(k, y) \geq 0$, where $y \in [0, 1]$ measures policy leniency. The case $y = 1$ corresponds to no intervention. With slight abuse of notation, let $V(k, \theta, 1) = V(k, \theta)$ and $\Pi(k, 1) = \Pi(k)$. In the polar case of $y = 0$, the firm is not permitted to produce, so $V(k, \theta, 0) = \Pi(k, 0) = 0$. We assume that the investor benefits from a more lenient policy, all else equal: $\Pi_y(k, y) > 0$ for intermediate levels of y. Absent any investment agreement, the host country chooses regulatory policy $Y^0(k, \theta) \in \arg\max_{y\in[0,1]} V(k, \theta, y)$ as a function of investment k and the regulatory shock θ.

It is not obvious how to define an appropriate benchmark in the carve-out compensation scheme, against which to compare the measure by the host country. We let jointly efficient regulation serve as our benchmark. If k^J is the level of investment that maximizes expected joint surplus, then $y^J(\theta) \in \arg\max_{y\in[0,1]}\{V(k^J, \theta, y) + \Pi(k^J, y)\}$ is the level of regulation that maximizes ex post joint surplus under efficient investment k^J and given the regulatory shock θ.

Proposition A.4 *The compensation scheme*

$$T(k, \theta, y) = \begin{cases} \max\{\Pi(k^J, y^J(\theta)) - \Pi(k^J, y); 0\} & \text{for } k = k^J \\ 0 & \text{for } k \neq k^J \end{cases} \tag{A.12}$$

can implement the jointly efficient outcome $(k^J, y^J(\theta))$ *under partial regulation if*

$$\pi^J \equiv \int_{\underline{\theta}}^{\bar{\theta}} \Pi(k^J, y^J(\theta))dF(\theta) - R(k^J) \geq \pi^0 \equiv \max_{k\geq 0}\{\int_{\underline{\theta}}^{\bar{\theta}} \Pi(k, Y^0(k, \theta))dF(\theta) - R(k)\}. \tag{A.13}$$

Proof: The ex post jointly efficient level of regulation under partial regulation is $Y^J(k, \theta) \in \arg\max_{y\in[0,1]}\{V(k, \theta, y) + \Pi(k, y)\}$. The jointly efficient investment solves

$$k^J \in \arg\max_{k\geq 0}\{\textstyle\int_{\underline{\theta}}^{\bar{\theta}}[V(k, \theta, Y^J(k, \theta)) + \Pi(k, Y^J(k, \theta))]dF(\theta) - R(k)\}.$$

Let $y^J(\theta) \equiv Y^J(k^J, \theta)$. Absent an agreement, the ex post optimal host country regulation is $Y^0(k, \theta) \in \arg\max_{y\in[0,1]} V(k, \theta, y)$. The firm invests:

$$k^0 \in \arg\max_{k\geq 0}\{\textstyle\int_{\underline{\theta}}^{\bar{\theta}}\Pi(k, Y^0(k, \theta))dF(\theta) - R(k)\}.$$

By $\Pi_y(k,y) > 0$ and the revealed preference for $Y^0(k,\theta)$:

$$V(k,\theta,y)+\Pi(k,y) < V(k,\theta,Y^0(k,\theta))+\Pi(k,Y^0(k,\theta)) \text{ for all } y\in[0,Y^0(k,\theta)),\ Y^0(k,\theta)>0.$$

Hence, $Y^J(k,\theta) \geq Y^0(k,\theta)$. If $Y^0(k,\theta)\in(0,1)$, then $Y^J(k,\theta) > Y^0(k,\theta)$ by

$$[V_y(k,\theta,y)+\Pi_y(k,\theta,y)]|_{y=Y^0(k,\theta)} = \Pi_y(k,\theta,y)|_{y=Y^0(k,\theta)} > 0.$$

Given (k,θ), the host country has an incentive to overregulate in the sense of $Y^0(k,\theta) < Y^J(k,\theta)$ unless $Y^0(k,\theta)=1$ or $Y^J(k,\theta)=0$.

Consider now an investment agreement with compensation scheme (A.12). Suppose the firm has invested k^J, and consider the ex post optimal regulation by the host country. Since $y^J(\theta)$ maximizes $V(k^J,\theta,y)+\Pi(k^J,y)$, then for all $y\in[0,y^J(\theta)]$:

$$\begin{aligned} V(k^J,\theta,y)-T(k^J,\theta,y) &= V(k^J,\theta,y)+\Pi(k^J,y)-\Pi(k^J,y^J(\theta)) \\ &\leq V(k^J,\theta,y^J(\theta))+\Pi(k^J,y^J(\theta))-\Pi(k^J,y^J(\theta)) = V(k^J,\theta,y^J(\theta)). \end{aligned}$$

Hence, the host country cannot benefit from overregulating, i.e. setting $y<y^J(\theta)$ if $y^J(\theta)>0$. If $y^J(\theta)<1$, then

$$\begin{aligned} V(k^J,\theta,y)-T(k^J,\theta,y) &= V(k^J,\theta,y) = V(k^J,\theta,y)+\Pi(k^J,y)-\Pi(k^J,y) \\ &\leq V(k^J,\theta,y^J(\theta))+\Pi(k^J,y^J(\theta))-\Pi(k^J,y) \\ &< V(k^J,\theta,y^J(\theta)) \end{aligned}$$

for all $y\in(y^J(\theta),1]$. Hence, the host country cannot benefit from underregulating either. We conclude that the host country optimally sets $y=y^J(\theta)$ under the compensation scheme $T(k,\theta,y)$ for all realizations of θ, if the firm has invested k^J.

Consider finally the firm's investment incentive. The firm earns expected profit π^J defined in (A.13) by investing $k=k^J$. If the firm deviates by setting $k\neq k^J$, then it effectively has no investment protection because $T(k,\theta,y)=0$. It then anticipates to face the regulatory intervention $Y^0(k,\theta)$, which implies expected deviation profit $\int_{\underline{\theta}}^{\bar{\theta}}\Pi(k,Y^0(k,\theta))dF(\theta)-R(k)\leq\pi^0$, where π^0 was also defined in (A.13). All deviations $k\neq k^J$ are unprofitable if $\pi^J\geq\pi^0$.■

Proposition A.4 shows that a compensation scheme (A.12) that either completely undoes all negative consequences of regulation (relative to the efficient benchmark) or offers no compensation at all, can implement the jointly efficient outcome under partial regulation ($y\in[0,1]$) under similar circumstances that a carve-out compensation policy can implement the jointly efficient outcome under complete regulation ($y\in\{0,1\}$). The sufficient condition (A.13) is similar to condition (7) for $H=1$. The extension of the proposition to multiple investors is straightforward if we assume that the operating profit of each firm is strictly increasing in y.

A scheme such as (A.12) cannot implement the Nash Bargaining Solution, contrary to what is achieved in the setting of Proposition 2. Under complete regulation and carve-out compensation, the only problem is overregulation. Under efficient investment, this incentive is corrected for all levels of investment protection satisfying $\theta^{JC} \geq \theta^{J}$. This property offers a degree of freedom which can be used to distribute surplus across the negotiating parties. Under partial regulation, the problem is both under- and overregulation, which narrowly pins down the efficient compensation scheme.

A.5.5 Investment subsidies

Assume that each foreign firm h receives an investment subsidy that is a function $Z^h(\mathbf{k})$ of the total portfolio $\mathbf{k}$ of foreign investment. Let $\mathbf{Z} \equiv (Z^1, ..., Z^H)$ be the subsidy scheme. The agreement now is a pair $(\mathbf{Z}, \mathbf{T})$ of subsidy and compensation schemes for investments undertaken by Foreign firms in Home. Assume that $\Pi^h(\mathbf{k})$ is strictly concave in k_h and that $R^h(k_h)$ is convex for all firms.

Under $(\mathbf{Z}, \mathbf{T})$, it is ex post optimal for the host country to allow production if and only if

$$V(\mathbf{k}, \theta) + \textstyle\sum_{h=1}^{H}[Z^h(\mathbf{k}) + T^h(\mathbf{k},\theta)] \geq \sum_{h=1}^{H} Z^h(\mathbf{k}).$$

This incentive and the threshold $\Theta^J(\mathbf{k})$ are both independent of $\mathbf{Z}$ because investment support is sunk when the host country decides whether to allow production or to regulate. The equilibrium investment $\hat{\mathbf{k}}$ under this agreement solves for all $h \in \mathcal{H}$:

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

where $\hat{\mathbf{k}}_{-h} = (\hat{k}_1, ..., \hat{k}_{h-1}, \hat{k}_{h+1}, ..., \hat{k}_H)$.

Followng the steps used in Appendix A.4 to prove Proposition 4, it is straightforward to verify the following (the proof is omitted):

Lemma 3 *For any agreement with investment support* $\mathbf{Z}$ *and general compensation* $\mathbf{T}$*, there exists an alternative agreement* $(\mathbf{Z}, \hat{\mathbf{T}})$ *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 (*$\tilde{\Pi}^h(\mathbf{Z}, \hat{\mathbf{T}}) = \tilde{\Pi}^h(\mathbf{Z}, \mathbf{T})$ *for all* h*), and gives the host country weakly higher expected surplus (*$\tilde{V}(\mathbf{Z}, \hat{\mathbf{T}}) \geq \tilde{V}(\mathbf{Z}, \mathbf{T})$*).*

The modified agreement yields expected investment profit

$$\tilde{\Pi}^h(\mathbf{Z}, \hat{\mathbf{T}}) \equiv Z^h(\hat{\mathbf{k}}) + F(\hat{\theta})\Pi^h(\hat{\mathbf{k}}) + \int_{\hat{\theta}}^{\bar{\theta}} \hat{T}^h(\hat{\mathbf{k}}, \theta)dF(\theta) - R^h(\hat{k}_h)$$

to firm h and expected host country surplus

$$\tilde{V}(\mathbf{Z},\hat{\mathbf{T}}) \equiv \int_{\underline{\theta}}^{\hat{\theta}} V(\hat{\mathbf{k}},\theta)dF(\theta) - \textstyle\sum_{h=1}^{H}[Z^h(\hat{\mathbf{k}}) + \displaystyle\int_{\hat{\theta}}^{\bar{\theta}} \hat{T}^h(\hat{\mathbf{k}},\theta)dF(\theta)],$$

where $\hat{\theta} \equiv \hat{\Theta}(\hat{\mathbf{k}})$ is the equilibrium threshold for regulation under $(\mathbf{Z},\hat{\mathbf{T}})$. Lemma 3 says nothing about the structure of compensation payments $\hat{\mathbf{T}}$. These are specified in the following result.

Lemma 4 *For any agreement with investment support* $\mathbf{Z}$ *and general compensation* $\hat{\mathbf{T}}$ *that yields a threshold for regulation* $\hat{\Theta}(\mathbf{k}) \in [\Theta(\mathbf{k}),\Theta^J(\mathbf{k})]$, *there exists an alternative agreement* $(\mathbf{Z}^C,\mathbf{T}^C)$ *where*

$$T^{hC}(\mathbf{k},\theta,\hat{\theta}) \equiv \begin{cases} \Pi^h(\mathbf{k}) & \text{if } \theta \le \hat{\theta} \\ 0 & \text{if } \theta > \hat{\theta} \end{cases} \quad \textit{for all } h \in \mathcal{H}, \tag{A.14}$$

that offers firms the same expected investment profit and gives the host country the same expected surplus as in $(\mathbf{Z},\hat{\mathbf{T}})$ $(\tilde{\Pi}^h(\mathbf{Z}^C,\mathbf{T}^C) = \tilde{\Pi}^h(\mathbf{Z},\hat{\mathbf{T}})$ *for all* h *and* $\tilde{V}(\mathbf{Z}^C,\mathbf{T}^C) = \tilde{V}(\mathbf{Z},\hat{\mathbf{T}}))$.

Proof: Define the subset $\boldsymbol{\kappa}(\hat{\theta}) = \{\mathbf{k} \ge \mathbf{0} : \Theta(\mathbf{k}) \le \hat{\theta}\}$ and its complement $\boldsymbol{\kappa}^c(\hat{\theta}) = \{\mathbf{k} \ge \mathbf{0} : \Theta(\mathbf{k}) > \hat{\theta}\}$, where $\hat{\theta} \equiv \hat{\Theta}(\hat{\mathbf{k}})$. In particular, equilibrium investment satisfies $\hat{\mathbf{k}} \in \boldsymbol{\kappa}(\hat{\theta})$. Consider an investment agreement $(\mathbf{Z}^C,\mathbf{T}^C)$ with support support scheme $\mathbf{Z}^C$ characterized by

$$Z^{hC}(\mathbf{k}) \equiv \begin{cases} Z^h(\hat{\mathbf{k}}) + \int_{\hat{\theta}}^{\bar{\theta}} \hat{T}^h(\hat{\mathbf{k}},\theta)dF(\theta) + [R_h^h(\hat{k}_h) - F(\hat{\theta})\Pi_h^h(\hat{\mathbf{k}})](k_h - \hat{k}_h) & \text{if } \mathbf{k} \in \boldsymbol{\kappa}(\hat{\theta}) \\ Z^h(\mathbf{k}) + \int_{\hat{\Theta}(\mathbf{k})}^{\bar{\theta}} \hat{T}^h(\mathbf{k},\theta)dF(\theta) & \text{if } \mathbf{k} \in \boldsymbol{\kappa}^c(\hat{\theta}) \end{cases} \quad \text{for all } h \in \mathcal{H} \tag{A.15}$$

and carve-out compensation $\mathbf{T}^C$, where $T^{hC}(\mathbf{k},\theta,\hat{\theta})$ was defined (A.14), and we have set $\hat{\theta} = \hat{\Theta}(\hat{\mathbf{k}})$.

Consider the profit-maximizing investment by firm h under $(\mathbf{Z}^C,\mathbf{T}^C)$ if all other firms have invested $\hat{\mathbf{k}}_{-h}$. All investments k_h such that $(k_h,\hat{\mathbf{k}}_{-h}) \in \boldsymbol{\kappa}(\hat{\theta})$, yield expected investment profit

$$\begin{aligned} & Z^{hC}(k_h,\hat{\mathbf{k}}_{-h}) + F(\hat{\theta})\Pi^h(k_h,\hat{\mathbf{k}}_{-h}) - R^h(k_h) \\ = \; & Z^h(\hat{\mathbf{k}}) + F(\hat{\theta})\Pi^h(k_h,\hat{\mathbf{k}}_{-h}) + \int_{\hat{\theta}}^{\bar{\theta}} \hat{T}^h(\hat{\mathbf{k}},\theta)dF(\theta) + [R_h^h(\hat{k}_h) - F(\hat{\theta})\Pi_h^h(\hat{\mathbf{k}})](k_h - \hat{k}_h) - R^h(k_h). \end{aligned}$$

Obviously, $k_h = \hat{k}_h$ is the profit-maximizing investment in $\boldsymbol{\kappa}(\hat{\theta})$. This investment yields expected profit

$$\tilde{\Pi}^h(\mathbf{Z}^C,\mathbf{T}^C) = Z^h(\hat{\mathbf{k}}) + F(\hat{\theta})\Pi^h(\hat{\mathbf{k}}) + \int_{\hat{\theta}}^{\bar{\theta}} \hat{T}^h(\hat{\mathbf{k}},\theta)dF(\theta) - R^h(\hat{k}_h) = \tilde{\Pi}^h(\mathbf{Z},\hat{\mathbf{T}}).$$

All deviations to $k_h \neq \hat{k}_h$ such that $(k_h, \hat{\mathbf{k}}_{-h}) \in \hat{\boldsymbol{\kappa}}^C(\hat{\theta})$ are unprofitable by:

$$
\begin{aligned}
& Z^{hC}(k_h, \hat{\mathbf{k}}_{-h}) + F(\max\{\hat{\theta}; \Theta(k_h, \hat{\mathbf{k}}_{-h})\})\Pi^h(k_h, \hat{\mathbf{k}}_{-h}) - R^h(k_h) \\
= \; & Z^h(k_h, \hat{\mathbf{k}}_{-h}) + F(\hat{\Theta}(k_h, \hat{\mathbf{k}}_{-h}))\Pi^h(k_h, \hat{\mathbf{k}}_{-h}) + \int_{\hat{\Theta}(k_h, \hat{\mathbf{k}}_{-h})}^{\hat{\theta}} \hat{T}^h(k_h, \hat{\mathbf{k}}_{-h}, \theta) dF(\theta) - R^h(k_h) \\
& -[F(\hat{\Theta}(k_h, \hat{\mathbf{k}}_{-h})) - F(\max\{\hat{\theta}; \Theta(k_h, \hat{\mathbf{k}}_{-h})\})]\Pi^h(k_h, \hat{\mathbf{k}}_{-h}) \\
\leq \; & Z^h(k_h, \hat{\mathbf{k}}_{-h}) + F(\hat{\Theta}(k_h, \hat{\mathbf{k}}_{-h}))\Pi^h(k_h, \hat{\mathbf{k}}_{-h}) + \int_{\hat{\Theta}(k_h, \hat{\mathbf{k}}_{-h})}^{\hat{\theta}} \hat{T}^h(k_h, \hat{\mathbf{k}}_{-h}, \theta) dF(\theta) - R^h(k_h) \\
\leq \; & \tilde{\Pi}^h(\mathbf{Z}, \hat{\mathbf{T}}) = \tilde{\Pi}^h(\mathbf{Z}^C, \mathbf{T}^C)
\end{aligned}
$$

These results establish $\hat{\mathbf{k}}$ as an equilibrium investment portfolio under $(\mathbf{Z}^C, \mathbf{T}^C)$. Moreover, all firms have the same expected investment profit under $(\mathbf{Z}^C, \mathbf{T}^C)$ as $(\mathbf{Z}, \hat{\mathbf{T}})$. The equilibrium threshold for regulation is the same under $(\mathbf{Z}^C, \mathbf{T}^C)$ as $(\mathbf{Z}, \hat{\mathbf{T}})$ and given by $\hat{\theta} = \hat{\Theta}(\hat{\mathbf{k}})$. The host-country expected surplus is the same under both agreements

$$
\begin{aligned}
\tilde{V}(\mathbf{Z}^C, \mathbf{T}^C) &= \int_{\underline{\theta}}^{\hat{\theta}} V(\hat{\mathbf{k}}, \theta) dF(\theta) + \textstyle\sum_{h=1}^H Z^{hC}(\hat{\mathbf{k}}) \\
&= \int_{\underline{\theta}}^{\hat{\theta}} V(\hat{\mathbf{k}}, \theta) dF(\theta) + \textstyle\sum_{h=1}^H [Z^h(\hat{\mathbf{k}}) + \int_{\hat{\theta}}^{\hat{\theta}} \hat{T}^h(\hat{\mathbf{k}}, \theta) dF(\theta)] = \tilde{V}(\mathbf{Z}, \hat{\mathbf{T}}).
\end{aligned}
$$

The host and the source country are hence both indifferent between $(\mathbf{Z}, \hat{\mathbf{T}})$ and $(\mathbf{Z}^C, \mathbf{T}^C)$.■

Combining Lemma 3 and Lemma 4 yields:

Proposition A.5 *For any agreement* $(\mathbf{Z}, \mathbf{T})$*, there exists an alternative agreement* $(\mathbf{Z}^C, \mathbf{T}^C)$*, where*

$$
T^{hC}(\mathbf{k}, \theta, \hat{\theta}) \equiv \begin{cases} \Pi^h(\mathbf{k}) & \text{if } \theta \leq \hat{\theta} \\ 0 & \text{if } \theta > \hat{\theta} \end{cases} \quad \text{for all } h \in \mathcal{H}, \tag{A.16}
$$

that offers firms the same expected profit as in the initial agreement ($\tilde{\Pi}^h(\mathbf{Z}^C, \mathbf{T}^C) = \tilde{\Pi}^h(\mathbf{Z}, \mathbf{T})$ *for all h), and gives the host country weakly higher expected surplus* ($\tilde{V}(\mathbf{Z}^C, \mathbf{T}^C) \geq \tilde{V}(\mathbf{Z}, \mathbf{T})$)*.*

If the agreement includes provisions for investment support, then the agreement can do no better than to include provisions for regulatory expropriation built on carve-out compensation. Moreover, this compensation scheme is particularly simple in that the threshold for compensation is the same for all firms and independent of investment. This does not hinge on any specific assumptions of bargaining format, only that the negotiated outcome is Pareto optimal from the viewpoint of the contracting parties. Investment support can be used both to incite investment and to distribute surplus across host country and investors. The only role of investment protection is then to prevent underregulation in equilibrium. This can easily be accomplished by way of the simple carve-out scheme (A.16).

A.5.6 Imperfect enforcement

Verifiability of the regulatory shock Let $Q(|\theta - \theta^{JC}|)$ be the probability that an arbitration court correctly asserts the sign of $\theta - \theta^{JC}$ if firms have invested $\mathbf{k}^J$. Q is then the likelihood that the court enforces the agreement subsequent to regulation. The court is more likely to make a correct judgement, the more θ differs from θ^{JC} by an assumption that Q is strictly increasing in $|\theta - \theta^{JC}|$ for $Q < 1$.

The host country will regulate for all $\theta > \theta^J$, regardless of 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 $Q(\theta^{JC} - \theta^J) = 1$. But the latter does not require perfect enforcement in all states of the world, i.e. $Q(0) = 1$, only that the precision of the court is sufficiently high 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.

For comparison, 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 $Q(0) = 1$. 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 $\theta < \theta^J$.

Verifiability of foregone operating profits Assume that there is a single investor with operating profit $\Pi(\beta, k) \geq 0$, where β is a productivity parameter, and investment cost $R(k) \geq 0$. Both functions are strictly increasing in their arguments. The size k of investment is common knowledge and verifiable. But β is observed only by the firm, making $\Pi(\beta, k)$ private information. We assume that investment takes place before the revelation of β so that k does not signal productivity.

Let the investment agreement specify compensation $T(k, \theta, \beta)$ if the firm has operating profit $\Pi(\beta, k)$. The firm self-reports some operating profit $\Pi(b, k)$, or equivalently b. The court then performs an audit that reveals the true profit with probability $\rho(b, \beta) < 1$, and uncovers no information with probability $1 - \rho(b, \beta)$. In the latter case, the tribunal accepts the investor's self-reported profit.

A firm with productivity β therefore has expected profit

$$\rho(b,\beta)T(k,\theta,\beta) + (1-\rho(b,\beta))T(k,\theta,b)$$

from reporting b to the tribunal. In comparison, a truthful report yields compensation $T(k,\theta,\beta)$.

It is better for the firm to truthfully report its productivity β rather than b if and only if

$$T(k,\theta,\beta) \geq \rho(b,\beta)T(k,\theta,\beta) + (1-\rho(b,\beta))T(k,\theta,b).$$

A firm with productivity b prefers to truthfully report its productivity b rather than misrepresent its productivity to β if and only if

$$T(k,\theta,b) \geq \rho(\beta,b)T(k,\theta,b) + (1-\rho(\beta,b))T(k,\theta,\beta).$$

Combining these inequalities yields the necessary incentive compatibility constraint $T(k,\theta,b) = T(k,\theta,\beta) = T(k,\theta)$. Hence, the compensation mechanisms cannot be used to elicit any information from the firm ex post. Instead, the carve-out scheme must build on some external benchmark, such as expected operating profit $E_\beta[\Pi(\beta,k)]$.