

Regulatory Takings and Environmental Regulation in NAFTA's Chapter 11*

Emma Aisbett,[†] Larry Karp,[‡] and Carol McAusland[§]

February 20, 2006

Abstract

NAFTA's investment treaty has led to several expropriation compensation claims from investors hurt by new environmental regulations. Expropriation clauses in international treaties solve post-investment moral hazard problems such as hold-ups. However, these clauses can interact with National Treatment clauses in a manner that hinders investment. A police powers carve-out from the definition of expropriation can be Pareto-improving and can increase the level of foreign investment.

Keywords: foreign direct investment, regulatory takings, expropriation, NAFTA, National Treatment, environment

JEL classification numbers K3, Q58, F21

*Our thanks to Jon Sonstelie, Perry Shapiro and Ted Frech for helpful suggestions, to Dimitrios Reppas for research assistance, and to the Institute on Global Conflict and Cooperation for financial support.

[†]Department of Agricultural and Resource Economics, 207 Giannini Hall, University of California, Berkeley CA 94720 aisbett@are.berkeley.edu

[‡]Department of Agricultural and Resource Economics, 207 Giannini Hall, University of California, Berkeley CA 94720 karp@are.berkeley.edu

[§]Department of Agricultural and Resource Economics, 2200 Symons Hall, University of Maryland, College Park MD 20742 cmcausland@arec.umd.edu

1 Introduction

Bilateral Investment Treaties (BITs) and multilateral investment agreements (including Chapter 11 of NAFTA, the North America Free Trade Agreement) seek to promote foreign direct investment by offering foreign investors increased security and transparency. These treaties require that hosts compensate investors in the event of expropriation or (possibly) regulatory actions that diminish the value of the investment. Foreign investors can bring their expropriation claims against the host country to an international tribunal, bypassing the host's domestic courts. The international tribunal decides whether the host's action constitutes a "regulatory taking," requiring compensation, or whether it falls within the host's "police powers carve-out" for policies designed to protect the public good, and is therefore not compensatory. The tribunal's criteria for a regulatory taking, and its standard of proof for a carve-out exemption, may differ from the host's; some policies that would be exempt from compensation under *domestic* law may be compensatory under the terms of the investment treaty. Thus the expropriation clause of treaties like NAFTA, together with its investor-to-state provision, gives foreign investors protection not available to domestic investors. The investment agreement therefore subjects the host to risks from international investment that do not arise with domestic investment.

In many cases, the need for regulation is known only after the investment is sunk. The existence of sunk costs combined with the fact that the host makes the regulatory decision creates an efficiency argument for the host to bear the costs of this regulation. This argument supports a narrow police powers carve-out. Contingent contracts or lump sum transfers that are made before the investment is undertaken could in principle offset the host's "regulatory risk" (defined as the risk of bearing the costs caused by regulation). For example, if the host can make the right to invest conditional on a side-payment from the investor, then the host can assure itself non-negative expected value from any project that it accepts. However practical difficulties limit the use of contingent contracts, and treaty provisions that promote non-discrimination—particularly the National Treatment and Most Favored Nation (MFN) clauses—restrict the feasibility of lump sum transfers. Thus, a narrow police powers carve-out may lead to an allocation of risk that affects both the distribution of the rents and the efficiency of foreign investments.

The National Treatment clause requires that foreign investors receive at least as good treatment as domestic investors "in like circumstances"; the MFN clause similarly guarantees investors treatment at least as good as that received by investors from any other foreign country. The international tribunal decides whether the foreign and domestic firms are indeed

in like circumstances. The treaties' investor-to-state provision and the use of international tribunals create an asymmetry between domestic and foreign investors, causing their circumstances to be inherently "unlike". Thus, there is a tension between the expropriation and the non-discrimination clauses. Individually, each of these clauses helps to solve a moral hazard problem, but collectively they might lead to inefficiencies.¹ The requirement of National Treatment and MFN makes it difficult for the host to impose stricter requirements on foreign investors in order to offset the increased risk created by the expropriation clause. Thus, the non-discrimination clauses create an efficiency argument for courts to adopt a narrow interpretation of regulatory taking, i.e. permit a broad police powers carve-out.

Expropriation clauses appear in almost all of the more than 2200 BITs and modern plurilateral investment treaties, including NAFTA, the Central American - Dominican Republic Free Trade Agreement (CAFTA-DR), and the draft Free Trade Agreement of the Americas (FTAA). The widespread use of expropriation clauses is controversial, not least because of the early experience with NAFTA. Interpreting "expropriation" broadly, investors have sued NAFTA governments for actions such as banning a polluting petrol additive (Methanex v. US) and refusing to permit a hazardous waste facility (Metalclad v. Mexico). These lawsuits have prompted some critics to claim that Chapter 11 benefits multinational corporations at the expense of states' sovereign rights, the environment, and the public good; see Public Citizen (2001) and Mann (2001).

The decisions offered by the tribunals adjudicating these NAFTA investment cases have not cleared the controversy. In *Metalclad v. Mexico*, the tribunal found expropriation to have occurred even though there was no "obvious benefit [to] the host state" (International Centre for Settlement of Investment Disputes 2000). Turk (2005) concludes that "[c]ompensation is required when a government actor deprives a foreign investor of his fundamental ownership rights..., regardless of the means, or reasons for the expropriation." Conversely, *Methanex v. US* tribunal declared governments exempt from paying compensation for bona fide regulations for the public good. Gaines (2002) and Soloway (2002), writing well before this ruling, state that most tribunals acknowledge an implicit police powers carve-out from the definition of expropriation, freeing states from the obligation to compensate foreign investors for injury arising

¹Several recent papers, including Bagwell and Staiger (2001a), Bagwell and Staiger (2001b), and Ederington (2001) analyze the role of National Treatment clauses in the GATT/WTO. Those papers, like ours, are concerned with the interplay of different parts of an agreement. However, those papers involve international trade rather than investment, and their results are driven by a terms-of-trade effect, which has no role in our analysis.

from non-discriminatory regulation for the public good. These contradictory rulings leave unsettled the question of how future tribunals will interpret expropriation clauses.

In this paper, we assess the efficiency, equity, and investment promotion impacts of a police powers carve-out. Since many of the NAFTA cases involve environmental regulations, our model uses an environmental setting; however the tension between expropriation and non-discrimination clauses arises for a variety of regulatory problems. Similarly, even though our modeling choices are guided by characteristics of the legal environment created by NAFTA, our results are applicable to the many hundreds of similar bilateral and multilateral investment agreements containing expropriation and non-discrimination clauses.

The rest of this Introduction discusses the role of expropriation clauses and their relation to other parts of investment agreements. We also review relevant literature. Section 2 of the paper discusses specific cases that have come before NAFTA tribunals as well as details of investment agreements; this information validates the two-period model introduced in Section 3. That section examines the equilibrium level of regulation occurring post-investment. Section 4 studies the investment stage, where we assess the efficiency and equity of a narrow versus broad police powers carve-out. Section 5 concludes.

Expropriation clauses help a host to make a credible commitment not to expropriate foreign investors' sunk assets, thereby helping to solve the hold-up problem. Grossman and Hart (1986) provide one of the first models of the hold-up problem; recent contributions include Noldeke and Schmidt (1995) and Schmitz (2001). Markusen (2001) discusses the costs and benefits, for developing countries seeking to attract foreign direct investment (FDI), of the credible commitment that expropriation clauses help create.

Advocates argue in favor of using investment treaties to protect investors' rights on the assumption that these treaties promote FDI. An emerging body of empirical literature, including Hallward-Driemeier (2003), Neumayer and Spess (2005) and Tobin and Rose-Ackerman (2004), finds little conclusive evidence that BITs do promote FDI. Huang (2005) provides mixed evidence concerning foreign firms' need for international protection. Compared to average domestic firms, foreign firms report lower levels of regulatory constraint, but foreign firms may be treated less well than politically connected domestic firms.

Even a host that is not tempted to expropriate the foreign investor may impose excessively strict regulations, simply because the cost of the regulations is external to the host. Reviewing the literature on Pollution Havens, Wilson (1997) labels a host's failure to internalize the cost

that local regulation imposes on inward FDI as a NIMBY (Not In My Backyard) outcome; see also Levinson (1997). From the investor's perspective, zealous regulation can be viewed as creeping expropriation or a regulatory taking. Expropriation clauses that force governments to compensate investors for costs arising from regulation can induce hosts to internalize costs, thereby eliminating excess regulation and promoting investment. Cost-internalization is the leading justification among legal scholars for the U.S. Fifth Amendment, which states that private property shall not be "taken for public purpose without just compensation" (Been and Beauvais 2003). Cost-internalization is likely to be even more important when dealing with foreign investors, who often repatriate profits and have less domestic political leverage.

The extent of protection that the Fifth Amendment provides property owners is a matter of debate, but thus far U.S. courts have rejected the doctrine of regulatory takings (a term often associated with Epstein (1985)). Some Chapter 11 plaintiff arguments (a few of which are summarized in the next section) appear to construe NAFTA as an endorsement of the doctrine of regulatory takings for international investments. These plaintiff arguments have prompted public complaints that the expropriation clause amounts to an end-run around U.S. and other nations' domestic law. Some observers believe regulatory takings claims such as *Methanex v. U.S.* are an unintended consequence of the expropriation clause. Daniel Price, the lead U.S. Chapter 11 negotiator, rejects this view, stating "The parties did not stumble into this [interpretation]....This was a carefully crafted definition" (quoted from Greider (2001)); see also Price (2000). The legal community is still debating whether the NAFTA case history validates fears that NAFTA establishes a regulatory takings doctrine; see Tschen (1999), Soloway (2002), Gaines (2002), Alvarez and Park (2003) and Turk (2005).

The main argument in favor of a narrow police powers carve-out is that it promotes investment. However, a narrow carve-out may lead to excessive investment. Blume, Rubinfeld, and Shapiro (1984) point out that court acceptance of the doctrine of regulatory taking provides investors free insurance against the loss induced by condemnation, thereby possibly leading to excessive investment. If international tribunals accept the doctrine of regulatory taking by adopting a narrow police powers carve-out, while domestic courts continue to reject the doctrine of regulatory taking, foreign firms implicitly receive subsidized insurance that is denied to domestic investors.

We show that a narrow police powers carve-out may also hinder investment. A narrow police powers carve-out transfers regulatory risk to the host, reducing the host's expected payoff

from investment. If the carve-out is very narrow, the host's expected payoff may be negative and the host will want to reject some projects outright. Most investment treaties contain only weak rights of establishment, allowing the host to reject undesirable projects. Under these treaties, a narrow carve-out may compel the host to prohibit investment in risky sectors, even though the investment may increase global welfare.

Some agreements like NAFTA, though, guarantee stronger rights to invest. We show that under a narrow police powers carve-out, agreements with strong rights to invest may cause a host to receive more foreign investment, but benefit less. For this type of investment agreement, a broader carve-out can increase the host's expected benefit from FDI, and may be desirable on distributional grounds.

2 Background and institutional context

This section provides detail about NAFTA's expropriation clause, some NAFTA investor-to-state cases, and NAFTA features that might limit efficient first-stage bargaining between hosts and investors.

2.1 Expropriation and Chapter 11

NAFTA's Article 1110 governs expropriation and compensation. This article states:

1. No Party may directly or indirectly nationalize or expropriate an investment of an investor of another Party in its territory or take a measure tantamount to nationalization or expropriation of such an investment ("expropriation"), except:

- (a) for a public purpose;
- (b) on a non-discriminatory basis;
- (c) in accordance with due process of law and Article 1105(1) [fair and equitable treatment]; and
- (d) on payment of compensation in accordance with paragraphs 2 through 6 [detailing the size and currency of compensation payments].

2. Compensation shall be equivalent to the fair market value of the expropriated investment immediately before the expropriation took place ("date of expropriation"), and shall not reflect any change in value occurring because the intended

expropriation had become known earlier. Valuation criteria shall include going concern value, asset value including declared tax value of tangible property, and other criteria, as appropriate, to determine fair market value.

This article appears to state that even if the host's actions are for a public purpose, non-discriminatory, and in accordance with due process, they are still subject to compensation requirements.

To date twenty seven publicly reported compensation claims have been filed under this clause. These claims are arbitrated by either the International Centre for Settlement of Investment Disputes (ICSID) or United Nations Commission on International Trade Law (UNCITRAL). Below we describe three of these claims, focusing on those features that guide our modeling choices in Section 3.

2.2 Non-discrimination rules

Like many other investment agreements, Chapter 11 specifies minimum treatment standards, areas in which this treatment must be afforded, and exceptions to the treatment standards. NAFTA's National Treatment standard is more comprehensive than most BITs. Article 1102: National Treatment states

Each Party shall accord to investors of another Party treatment no less favorable than that it accords, in like circumstances, to its own investors with respect to the establishment, acquisition, expansion, management, conduct, operation, and sale or other disposition of investments.

The second paragraph is identical, with the phrase "to investors" replaced by "to investments of investors". Article 1103: Most-Favored-Nation Treatment is identical to the first two paragraphs of Article 1102 with the phrase "to its own investors" replaced by the phrase "to investors of any other Party or of a non-Party".

Articles 1102 and 1103 state that the host's equitable treatment obligations begin even before ground is broken for a new project, since the non-discrimination rules apply during establishment and acquisition. Thus, the rules prevent the host nation from demanding that foreigners pay up-front transfers in order to secure the right to invest.

Annexes I and II of NAFTA allow governments to select sectors in which they reserve the right not to provide National Treatment or MFN Treatment. All three NAFTA signatories took

advantage of these clauses, exempting industries such as atomic energy, the extraction, transport or processing of petroleum, gas, uranium and certain other minerals, aviation, shipping transport, trucking, telecommunications, media ownership, social services provision (e.g. law enforcement and correctional services), legal services, minority affairs, education, and fisheries.

2.3 Case Summaries

Ethyl Corporation v. Canada (See Canada Department of Foreign Affairs (2004), Environment Canada (1998), Warner (2005), and Global Lead Network (2005).) Ethyl Corporation claimed that the Canadian federal government's ban on imports and interprovincial transport of the gasoline additive MMT violated NAFTA articles 1102 - National Treatment, 1106 - Performance Requirements, and 1110 - Expropriation and Compensation. Ethyl emphasized that the ban would benefit producers of competing oxygenates (ethanol and MTBE), and that it was possible for MMT to be used throughout Canada if manufacturing plants were established in each province. On this basis Ethyl claimed that the ban was arbitrary and intentionally discriminatory toward Ethyl, which was the sole supplier of MMT in Canada.

The Canadian government claimed that the ban was motivated by health risks associated with manganese emissions as well as air quality concerns due to damage MMT might cause to the catalytic converters in cars. Canada also noted that MMT use in unleaded petrol was banned in much of the U.S. as well as in parts of Europe. (The U.S. EPA initially banned the sale of MMT but was forced to allow it by a domestic court ruling.) Canada's ban of import and transport of MMT, rather than *use*, weakened the government's position. Allegedly, Canada's own laws prevented it from banning MMT use because the responsible government agencies had found insufficient evidence of either the claimed environmental or health effects.

Ethyl submitted a claim for \$251 million to an UNCITRAL panel. Anticipating that the panel was likely to find against it, in July 1998 the Canadian government reversed the ban on MMT import and transport, paid Ethyl \$13 million in legal fees and damages, and signed a letter stating that MMT is not known to be hazardous and does not impair motor vehicle function.

Metalclad Corporation v. United Mexican States (See U.S Department of State (2005b) and British Columbia Supreme Court (2002).) The Metalclad Corporation, a U.S. waste disposal company, claimed that Mexico breached NAFTA articles relating to expropriation and standards of treatment. It asserted that a municipal government in Mexico wrongfully refused to permit Metalclad's subsidiary to open and operate a hazardous waste facility that Metalclad

had built. Metalclad claimed the project was built in response to the invitation of Mexican officials and that the project met all Mexican legal requirements. Metalclad further claimed that an ecological decree, made by the State government after Metalclad filed its initial claim against Mexico, constituted expropriation. The ecological decree made the area containing the landfill a preserve for cacti, preventing the landfill from ever being able to operate. Metalclad sought damages of US\$43,125,000 “plus damages for the value of the enterprise taken.”

In August 2000, the ICSID tribunal concluded that the Mexican government’s actions constituted expropriation and awarded Metalclad \$16.7 million. Discussing acts that constitute expropriation, in its statement of award, the tribunal writes “[t]hus, expropriation under NAFTA includes not only open, deliberate and acknowledged takings of property, such as outright seizure or formal or obligatory transfer of title in favour of the host State, but also covert or incidental interference with the use of property which has the effect of depriving the owner, in whole or in significant part, of the use or reasonably-to-be-expected economic benefit of property even if not necessarily to the obvious benefit of the host State.” (International Centre for Settlement of Investment Disputes (2000) para. 103) Although the British Columbia Supreme Court later set aside part of the award, it upheld the NAFTA tribunal’s finding that the State government’s ecological decree constituted an indirect expropriation.

The \$16.7 million award reflected Metalclad’s assessment of its costs for developing the landfill, not the value of the operation. This award amount appears inconsistent with language in NAFTA article 1110, paragraph 2, which states that compensation shall include asset value (which incorporates the present value of the stream of expected future profits). Another notable feature of the Metalclad case is that the Mexican government was held liable for the actions of its political subdivisions: the municipal agency charged with granting use permits and the State governor who declared the ecological decree.

Methanex Corporation v. United States of America (See U.S. Department of State (2005) and UNCITRAL Tribunal Methanex Corp. v. United States of America (2005).) Methanex Corporation, a Canadian marketer and distributor of methanol, alleged injuries resulting from a California ban on the use or sale in California of the gasoline additive MTBE, which uses Methanol as an ingredient. Methanex contended that a California Executive Order and regulations banning MTBE expropriated parts of its investments in the United States in violation of Article 1110, denied it fair and equitable treatment in accordance with international law in violation of Article 1105, and denied it National Treatment in violation of Article 1102. Cali-

California claimed that the ban on MTBE was enacted in order to protect public health and prevent water pollution. Methanex asserted that the Californian action resulted from political lobbying by the U.S. domestic ethanol industry, a competitor in the gasoline additive market. Methanex also contended that the regulations were not justified on environmental grounds and that there existed less drastic means of addressing the environmental issues.

Methanex claimed damages of \$970 million. On August 3, 2005, the Tribunal issued a Final Award dismissing all of Methanex's claims. The Tribunal dismissed the expropriation claim on the grounds that the MTBE ban was the result of due process and for the public good: "...as a matter of general international law, a non-discriminatory regulation for a public purpose, which is enacted in accordance with due process and, which affects, inter alia, a foreign investor or investment is not deemed expropriatory and compensable..." (UNCITRAL Tribunal, p. 278). This language indicates that a police powers carve-out is consistent with NAFTA's Chapter 11, at least in the eyes of this particular tribunal.² The Tribunal also ordered Methanex to pay the United States' legal fees and arbitral expenses of approximately \$4 million.

Glamis Gold Ltd. v. United States of America (See U.S Department of State (2005a), Glamis Gold Ltd. (2003) and United States of America (2005).) Glamis Gold Ltd., a publicly-held Canadian corporation engaged in the mining of precious metals, submitted a claim to arbitration for alleged injuries relating to a proposed gold mine in Imperial County, California. Glamis claims that it invested approximately \$15million in gold mining operations, including the purchase of mining claims for its Imperial County mining project. It asserts that \$13 of the \$15 million was spent after receiving initially favorable signals from various regulatory bodies. It further claims that certain federal government actions and California measures with respect to open-pit mining operations resulted in the expropriation of its investments in violation of Article 1110, and denied its investments the minimum standard of treatment under international law in violation of Article 1105. The California measures include regulations requiring backfilling and grading for mining operations in the vicinity of Native American sacred sites. The regulations are general and not specific to Glamis's operations. Nonetheless, seeking to respond to growing public opposition to the Glamis Imperial project, the then Governor Grey Davis explicitly stated that the regulations would essentially stop the project. Glamis claims damages

²The UNCITRAL panel's need not, however, create a precedent by which all future expropriation claims will be judged. The practice of *Stare Decisis* (letting precedent determine future legal rulings), is not followed in international law to the same extent as in the courts of some countries, notably the United States (Bhala 1999), (Chua 1998).

of not less than \$50 million, which it states is below the projected value of the investment. The U.S. defense claims that even without the actions of the Californian government, the project was uneconomic.

The case has not yet gone before the Tribunal.

3 The model

We develop a two-stage model of environmental regulation and investment in the presence of a pre-existing international investment agreement (IIA). In the first stage the host and the investor determine the size of T , the tax payment the source will make to the host conditional on the host doing nothing later that decreases the value of the project for the investor. The parties take as given the terms of the IIA, including the size of any police powers carve-out, and any other investment clauses such as National Treatment. The host's ability to reject a project may be constrained by a right to establishment requirement in the treaty; the investor has no such constraint.

The project might result in environmental damages of fixed magnitude H , a measure of a project's riskiness. In the first stage, the host and the investor have identical information concerning the probability distribution for this random event, and both agents are risk neutral. In the second stage the host obtains information about the probability of environmental harm; this information is not verifiable by the court.³ The host has three choices: (a) do not regulate, and suffer the possible environmental harm; (b) regulate, thereby avoiding the environmental harm, and reimburse the investor for the amount of the investment; or (c) regulate (avoiding the possible harm) and claim the police powers justification. The third action triggers an investor complaint to the international tribunal. Defending against this complaint is always costly to the host. We assume that if the host loses the litigation, it is unable to recontract with the investor.

³A multiperiod generalization of this model would allow the host to decide when to regulate. Even in this case it would, in general, not be optimal for the host to wait until it had perfect information about the environmental harm. Our model captures in a simple manner the idea that the host sometimes wants to impose regulations even before it is certain about damages in the unregulated state.

3.1 The second stage decision problem for the host

The first stage agreement stipulates that the host receives T in present discounted value of tax revenue, provided the host does nothing to diminish the investor's payoff. In the second period, the host decides whether to regulate. The purpose of the regulation might be to achieve an environmental objective that falls within the scope of the police powers carve-out, or it might be a means of disguised expropriation.

The standard definition of expropriation in international law requires that the host gain something from the allegedly expropriatory action. We define R as the expected value of *residual* benefits, excluding compensation payments, to the host from regulation. R is observable only to the host. In a traditional expropriation case, R represents the value to the host of the plant or property taken from the foreign investor. However, in the context of regulatory takings, it is a matter of debate whether the host must directly benefit from an action before that action can be labeled expropriation. Indeed, it is arguable whether any of the host governments actually benefited directly from first allowing and then expropriating the investments in the cases described in Section 2. The Glamis Gold case provides probably the best example of government gain. One could argue that, through the permitting process, U.S. governments at various levels sold to Glamis the right to remove the gold from U.S. soil. Then the Californian backfilling regulations took that right away. The end result was that the U.S. retained both the gold and the money from the sale of the mining claims. In light of the uncertainty whether a host must gain in order for an action to be deemed expropriation, we include R in our model, but none of our results rely on R being positive. When $R > T$ the host has an incentive to expropriate, creating a hold-up problem. If instead R is small, there is no holdup problem; however the lack of cost-internalization may still induce excessive regulation and therefore be globally inefficient.

In the interest of simplicity, we assume the investor loses the entire value of the investment if regulated.⁴ The investor's project cost, including the normal risk-adjusted return, is K . Conditional on the host not regulating, the present discounted value of the project in the second

⁴Whether the regulation reduces to 0 or merely diminishes the value of the project to the investor may be germane to the court's decision. If this is the case, then the host may need to decide on the form of regulation, not merely on whether to regulate. We ignore this kind of complication. For example, the host may have a binary choice, e.g., whether to allow the citing of a hazardous waste facility (as in *Metalclad v. United Mexican States*) or whether to allow the use of MTBE (as in *Methanex v. U.S.*).

stage is a random variable with expected value π . The expected value to the investor of the unregulated project is $\pi - T$. The value of π is learned in the second stage, and is not verifiable by the court, so the court cannot use it to calculate compensation. The investor can run the project more efficiently than the host ($\pi > R$), so transfer of ownership is never efficient for society.

It is common knowledge that the investment might cause environmental harm of magnitude H . After the investment has been made, the host learns β , the probability that this harm will actually occur ($0 \leq \beta \leq 1$); β is private information, and is uncorrelated with the stream of future profits, π . After learning β , the host takes one of the three actions described above.

If the host decides to regulate without paying compensation, the investor files a complaint. The tribunal imposes a test to determine whether the regulation falls within the scope of the police powers carve-out. Following the actual practice of NAFTA tribunals, e.g. in *Metalclad v. United Mexican States*, we assume the host must pay the investor compensation equal to K if the host fails the test. We assume that $\pi - T - K > 0$, i.e. the source's expected net return conditional on π and no regulation is positive.

In our setting the host has an unobservable incentive to use bogus environmental regulation as an indirect means of expropriation. The tribunal must use public details of the case to determine whether the regulation falls within the scope of the police powers carve-out, and is thus non-compensable. We assume that the tribunal's likeliness to accept the police powers carve-out defense is greater the higher is β . That is, the stronger the private evidence of environmental risk, the more likely that the public record and the tribunal acknowledge this risk. We also assume that, for a given realization of β , the tribunal is less likely to rule in favor of the host the more narrowly the IIA defines the police powers carve-out. Let N measure the narrowness of the police powers carve-out; N is set exogenously by the terms of the IIA and by the body of international law. A larger value of N corresponds to a narrower police powers carve-out, i.e. a stricter burden of proof for the host. We formalize these concepts via the function $P(\beta, N)$, defined as the probability the tribunal rules in favor of the host; the above assumptions imply $\partial P / \partial \beta > 0$ while $\partial P / \partial N < 0$.

In order to illustrate our main results concerning the efficient choice of N (in the following section), we use closed form expressions for the slopes of indifference curves. For this purpose,

we adopt the specification⁵

$$P(\beta, N) = \exp(- (1 - \beta) N). \quad (1)$$

The parameter N measures the narrowness of the police powers carve-out: for $N = 0$ the host always wins arbitration, so there is no investor protection; as $N \rightarrow \infty$ the host never wins, except in the limiting case where the host is certain that the project is harmful ($\beta = 1$).

The host always pays court costs C , which we take to be a constant. (Allowing C to be a non-decreasing function of N leads to uninteresting complications.) For simplicity, and to reflect the view that the variable legal costs are more onerous for the host, we ignore the investor’s court costs.⁶

We assume that K is greater than each of C and R . The inequality $K > C$ means that the host will litigate rather than voluntarily compensate the source if the host is certain of victory (i.e. if $P = 1$). $K > R$ implies the host would not make the investment on its own, since the cost of the investment (K) exceeds the payoff of the host-run investment (R).

3.2 The second stage outcome

The host’s and source’s expected payoffs under the three possible host actions (associated with the letters “ a, b, c ” respectively) are:⁷

⁵We also studied a model with the following binomial structure. The tribunal accepts the police powers carve-out defense if and only if the host’s regulation passes N independent and simultaneous “trials.” For example, the tribunal will simultaneously review whether the regulation was properly vetted with the scientific community, whether regulators were subject to undue influence from lobby groups, whether the investor was kept abreast of the policy process, and so on. Each trial might result in “pass” or “fail,” i.e. the regulation was properly vetted or not. Let p be the probability that a trial results in pass when the project is harmful, and $q \in (0, p)$ be the probability that the trial results in pass when the project is not harmful. The host’s subjective probability that the project is harmful is β , so its subjective probability of passing the test is $P(\beta, N) = \beta p^N + (1 - \beta) q^N$. This alternative formulation leads to the same qualitative results as described in this section of the paper. However, this alternative does not produce closed form expressions for the slopes of indifference curves, needed for the following section.

⁶In practice, tribunals often make the loser pay everyone’s court costs, as in *Methanex v. United States*. With this practice, there would be a fourth possible outcome to the second stage if the investor observes β ; the investor would accept regulation without compensation whenever $\beta > 1 + \frac{\ln \frac{K}{K+C}}{N}$. Assuming that the loser pays C would change details of the second stage equilibrium, but we have verified that there would be no qualitative changes in the first stage results. Accordingly, we maintain the simpler assumption that the host pays C .

⁷FDI may also generate positive spillovers for the host above and beyond tax payments. Higher employment and technology transfer are common examples. Adding a positive term to the host’s payoff under action (a) would

Host's Action	Host's 2nd stage expected payoff	Source's 2nd stage expected payoff
(a) do not regulate	$T - \beta H$	$\pi - T$
(b) regulate & compensate	$R - K$	K
(c) regulate & litigate	$R - (1 - P(\beta, N)) K$ $-C$	$(1 - P(\beta, N)) K$

Table 1: Expected payoffs and actions

Pair-wise comparison of host's payoffs under the three actions generates three threshold values:

$$\beta_1 \equiv \frac{N - \ln \frac{K}{C}}{N},$$

$$\beta_2 = z^{-1}(T) \text{ where } z(\beta) \equiv R - (1 - P(\beta, N))K - C + \beta H,$$

$$\beta_3 \equiv \frac{T + K - R}{H}.$$

For example, when the realization of β is greater than threshold value β_1 , the host will prefer to regulate and go to arbitration rather than to regulate and settle directly.

The value of T at which $\beta_1 = \beta_2 = \beta_3$ is

$$\tilde{T} \equiv H + R - K + \frac{H}{N} \ln \frac{C}{K}. \quad (2)$$

\tilde{T} is a threshold tax level beyond which the host is unwilling to regulate and settle directly. For a given realization of β , the higher is the tax level the greater the host's expected payoff from *not* regulating. Once T exceeds the threshold value \tilde{T} , the host will never forgo its tax revenues in favor of regulation unless it also thinks that the probability of harm is high enough to make defending its case before a tribunal preferable to paying compensation outright.

The following proposition uses the above definitions to identify the host's optimal second stage actions contingent on the realization of β . (All proofs are in Appendix A.)

capture this feature.

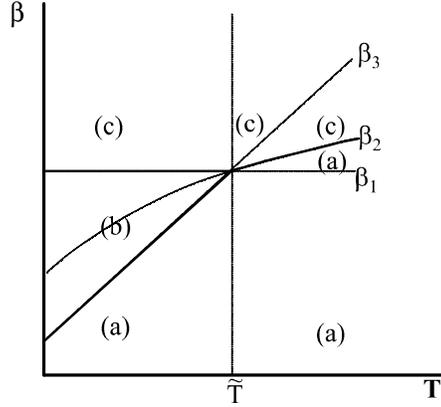


Figure 1: Second period Host actions, for constant N

Proposition 1 *In the regulation stage (second period), the host's optimal action depends on the realization of β .*

- *If $T \in [0, \tilde{T})$ then $\beta_3 < \beta_2 < \beta_1$ and the host will: regulate and litigate if $\beta > \beta_1$; regulate and directly compensate if $\beta \in (\beta_3, \beta_1)$; not regulate if $\beta < \beta_3$.*
- *If $T \geq \tilde{T}$ then $\beta_1 \leq \beta_2 \leq \beta_3$ and the host will: regulate and litigate if $\beta > \beta_2$; not regulate if $\beta < \beta_2$; never directly compensate.*

Figure 1 illustrates the relationships listed in Proposition 1 for the case $\tilde{T} > 0$; the letter a , b , or c in each of the Figure's regions identifies the host's optimal action, defined in Table 1. The graphs of β_2 and β_3 are increasing in T , while β_1 is independent of T , and the three graphs intersect at the same point. $\tilde{T} > 0$ if and only if

$$N > \frac{H}{H + R - K} \ln \frac{K}{C}. \quad (3)$$

Inequality (3) shows that the judicial test must be sufficiently strict in order for the host to prefer paying compensation rather than litigating, given that it is optimal to regulate.

Comparative statics of an increase in N For N sufficiently close to 0 (where β_1 and \tilde{T} are negative), the court almost always finds in favor of the host. In this case the host regulates for all values of β whenever its tax revenue, T , is less than the benefit of regulation minus the litigation costs, $R - C$. For large β this regulation achieves an environmental goal, and for

small β the regulation is merely disguised expropriation. The absence of a strict burden of proof for the police powers carve-out creates a hold-up problem at the investment stage, because the investor knows that the host will have an incentive to expropriate at a later stage.

If $\tilde{T} > 0$, as in Figure 1, an increase in N decreases the region (labelled ‘c’) in (T, β) space where the host decides to litigate; the increase in N expands both the region in which the host decides to regulate and compensate (‘b’), and the region in which the host decides not to regulate (‘a’).

For simplicity, we assume hereafter that the probability distribution for β does not have a mass point at $\beta = 1$. With this assumption, the probability that the host will litigate when $N = \infty$ is 0. As N becomes large, the host has negligible chance of prevailing in court. As $N \rightarrow \infty$, $\beta_1 \rightarrow 1$. In this limiting case, corresponding to no police powers carve-out, the host pays compensation if $\beta > \beta_3$, and otherwise refrains from regulating.

Social inefficiency There are two sources of inefficiency at this stage of the model. The first is that the host sometimes litigates, incurring socially wasteful court costs C . If this were the only source of inefficiency, it could be made negligible by choosing N sufficiently large, so that the host never litigates.

However, even when $N = \infty$, inefficiency persists if the required compensation, K , is less than the investor’s lost profits, $\pi - T$. In this case, the host does not fully compensate the investor, and therefore does not fully internalize the cost of regulation; this is true even if R is zero. The host regulates if and only if $R - K > T - \beta H$, i.e. when $\beta H > T + K - R$. In contrast, the socially efficient program requires regulation if and only if joint welfare with regulation, R , exceeds joint welfare without regulation, $\pi - \beta H$. Regulation is efficient ex post if and only if $\beta H > \pi - R$. Provided that the host has positive expected profits ($\pi > T + K$), the socially optimal minimal value of β that triggers regulation is strictly higher than the value of β above which the host regulates. In this sense, the host regulates too often.

Second period equity For N sufficiently large, there are combinations of β and T for which the host’s second-period expected payoff is negative. In these cases, the host is worse off than if it had never accepted the project. Equity concerns may make this kind of outcome unattractive. A smaller value of N (i.e., a broader police powers carve-out) would address these concerns.

4 The investment stage

During the first stage, the host and source determine the size of the tax payment T and the investor decides whether to undertake the investment project. Depending on the details of the IIA, the host may also be able to veto the project. In this stage, actors take the breadth of the police powers carve-out as pre-determined. They also take as given other clauses in the IIA, notably National Treatment and rules granting right of establishment.

An outcome to the first stage is a pair (T, N) . We examine the equity, efficiency and investment implications of a variety of outcomes. In particular, we show that IIAs with high N may discourage investment and may be inequitable. We begin by calculating each agent's first stage (ex ante) expected welfare conditional on the project proceeding. We then turn to the equity and efficiency of an outcome. A numerical example closes the section.

4.1 Ex ante expected welfare

The channel through which the host extracts T from foreign investors merits further discussion. If the IIA allows the host to demand side payments for access to its market, then T would include the value of those payments, in addition to the present value of future taxes. An IIA that contains a right of establishment clause prohibits these side payments, making it more difficult for the host to extract rent from the investor in the pre-investment stage. This limit on the host's bargaining power benefits the investor. The host might also benefit from this prohibition, because it decreases the scope for bribing corrupt officials (Markusen 2001).

We assume that the host is not able to demand a side payment as a condition for investment. Thus, T consists only of revenues collected from foreign-owned enterprises during general taxation. This limitation is the key to our results. National Treatment clauses stipulate that governments may not treat the investments of foreigners less favorably than those of citizens. In terms of general profit taxes, governments may not charge foreigners a higher tax rate. With this in mind, let t^d define the host's optimal *domestic* tax rate. If the tax rate t necessary to deliver T from the potential project is less than t^d , then the National Treatment clause does not prevent the host from obtaining T .⁸ However, if $t > t^d$ the host cannot obtain T without violating

⁸National treatment standards in international investment agreements do not prevent hosts from providing more favorable treatment to foreigners. A substantial literature addresses the potentially harmful tax competition between countries trying to attract FDI (de Mooij and Ederveen 2003). Concerns about such competition have led to criticism of the one-sided nature of national treatment clauses in investment agreements (Panagariya 1999).

the National Treatment clause, unless it increases the domestic tax rate above its optimal level. Standard tax considerations and the theory of the excess burden of taxation explain why raising taxes above t^d imposes costs on the host, regardless of whether the host eventually regulates the foreign-owned project.⁹

We use a reduced form model of the excess burden from domestic taxation. We assume that domestic and foreign investment markets are independent, and denote the deadweight loss from excessive taxation in the domestic market by $D(T)$.^{10, 11} The amount of revenue generated by the tax rate t^d is denoted T^d . By definition of deadweight loss, $D(T) > 0$ for $T > T^d$. Thus the host's benefit from payment T is $T - D(T)$. A positive value of $D(T)$ implies a first stage inefficiency: some of the rents transferred from the investor to the host are dissipated via tax distortions in the domestic investment market. The requirement of National Treatment caps the payment T the host can extract from the source without paying a penalty arising from distortions in its own market. This constraint may limit the ability of a larger T to offset the risk associated with higher N , causing the first stage outcome to be inefficient.

We now describe the host and foreign investor's ex ante expected welfare as a function of (T, N) . Both agents have rational expectations regarding the host's second-stage decision; they understand how this decision depends on (T, N) and the realization of β . Both agents' subjective probability distribution for β is uniform over $[0, 1]$. The discount factor is set equal to one and we assume that π is independent of β .

Some investor-to-state disputes are settled even before arbitration is launched (UNCTAD 2005). We therefore restrict our analysis to the most interesting case, where there are realizations of β for which the host wants to take any of the three possible second-stage actions. From equation (2), this possibility requires that $T < \tilde{T}$. We define an additional variable, $\bar{T} \equiv \frac{R+\pi}{2} - K$, whose role will be apparent in Lemma 1 below. In the interest of brevity we restrict our attention to taxes satisfying $T \in (\bar{T}, \tilde{T})$. This interval is independent of the deadweight loss function $D(T)$. This fact is important because it means that there always exists a

⁹If the elasticities of supply of foreign and domestic investment differ, the host has an economic reason for imposing different taxes on the two types of investment; there may also be positive externalities associated with domestic investment. In general, the optimal tax (from the standpoint of the host) levied on domestic investment may be lower than the optimal tax levied on foreign investment.

¹⁰The MFN principle might similarly generate deadweight loss. The host's share of any such loss can be incorporated into $D(T)$.

¹¹The host can use T to reduce the need for domestic taxation, so in a full model of international taxation, t^d and T are not independent. In order to keep the model simple, we ignore these kinds of interactions.

function $D(T)$ for which the conditions in the lemmas and propositions below are satisfied, so the results are not vacuous.

Table 1 above gives the host's and source's second period expected payoffs conditional on the realization of β . Using the entries in that table, the ex ante expected payoff, $E(W_H)$, for the host from accepting the project is

$$E(W_H) = \int_0^{\beta_3} (T - \beta H) d\beta + \int_{\beta_3}^{\beta_1} (R - K) d\beta + \int_{\beta_1}^1 (R - (1 - P(\beta, N)) K - C) d\beta - D(T), \quad (4)$$

and the ex ante expected payoff, $E(W_S)$, for the source from investing K is

$$E(W_S) = \int_0^{\beta_3} (\pi - T) d\beta + \int_{\beta_3}^{\beta_1} K d\beta + \int_{\beta_1}^1 (1 - P(\beta, N)) K d\beta - K. \quad (5)$$

4.2 Efficiency

In order to determine the shape of the respective indifference curves in (T, N) space, we first collect the partial derivatives of $E(W_H)$ and $E(W_S)$ in Table 2, using the definition $\alpha = 1 + \ln K - \ln C$.

$\frac{\partial(E(W_H))}{\partial T} = \frac{K+T-R}{H} - D'$	$\frac{\partial(E(W_H))}{\partial N} = \frac{\alpha C - K}{N^2}$
$\frac{\partial(E(W_S))}{\partial T} = -\frac{(K-R+T) - (\pi - K - T)}{H}$	$\frac{\partial(E(W_S))}{\partial N} = \frac{K - C}{N^2}$

Table 2. Derivatives of payoffs

Using these derivatives, the slope of the host's indifference curve is

$$\left. \frac{dN}{dT} \right|_{dE(W_H)=0} = \left[\frac{K - R + T - D'(T)H}{K - \alpha C} \right] \frac{N^2}{H}, \quad (6)$$

and the slope of the source's indifference curve is

$$\left. \frac{dN}{dT} \right|_{dE(W_S)=0} = \left[\frac{K - R + T - (\pi - T - K)}{K - C} \right] \frac{N^2}{H}. \quad (7)$$

The shape of the indifference curves depends on the following conditions.

C1.	$0 < T < \tilde{T}$
C2.	$D'(T) < \frac{K+T-R}{H}$
C3.	$T > \bar{T}$ where $\bar{T} \equiv \frac{R+\pi}{2} - K$
C4.	$D''(T) > \frac{1}{H} + \frac{2N}{K-\alpha C} \left(\frac{T+K-R}{H} - D'(T) \right)^2$

Table 3. Conditions on T and $D(\cdot)$

Lemma 1 *When conditions C1, C2, and C3 hold, both the host and the source's indifference curves are upward sloping in T, N space.*

Conditions C2 and C3 should hold in any outcome. Condition C2 ensures the marginal domestic distortion from T is not so large that it makes raising T immediately harmful to the host. (Section 4.3 discuss problems arising when T violates C2.) Condition C3 ensures that higher taxes do not benefit the source (by reducing the risk of regulation). Under conditions C1-C3, each indifference curve has a positive slope, and a north-west movement (in the T, N plane) increases the source's welfare and reduces the host's welfare.

Next we consider the convexity of agents' preferences.

Lemma 2 *The source has convex preferences over T and N when conditions C1 and C3 hold. The host has convex preferences over T and N when conditions C1, C2, and C4 hold.*

Convexity of the host's preferred set is equivalent to concavity of the host's indifference curve (N as a function of T). Convexity of the host's preferred set requires that D'' is sufficiently large.

Lemmas 1 and 2 are pivotal for understanding the efficiency of a broad versus narrow police powers carve-out. We start with the case where D' and D'' are negligible, i.e. where the host bears little if any cost when extracting payment T from the source. Supporters of narrow carve-outs likely have this scenario in mind.

Proposition 2 *If conditions C1, C2 and C3 hold, if $\pi - T - K \geq 0$, and if the tax distortion is sufficiently small (i.e. $D' \approx D'' \approx 0$), then, for any N^o, T^o pair satisfying $T^o \in (\bar{T}, \tilde{T})$, there exists a Pareto superior pair N', T' where $N' > N^o$ and $T' > T^o$.*

When $D'' \approx 0$, condition C4 is violated and the host's preference set is not convex. Moreover, equations (6) and (7) imply that for $D' \approx 0$ the host's indifference curve is steeper than the source's. In this case, any outcome (T, N) with $T < \tilde{T}$ and finite N is inefficient: the host and source would both benefit from a narrower carve-out and a correspondingly larger transfer T to the host. Proposition (2) assumes that $T < \tilde{T}$. Appendix B shows that outcomes with finite N are also inefficient when $T \geq \tilde{T}$ if $D' \approx 0$. These results imply that a (vanishingly) narrow police powers carve-out is necessary for efficiency when $D' \approx D'' \approx 0$. That is, when there is only negligible inefficiency at the investment stage, a broad police powers carve-out is inefficient. In our model with symmetric first-stage information and risk neutral agents, this leaves only the second-stage inefficiencies—the hold-up and cost-internalization problems—for which a strict rule regarding expropriation and compensation is an appropriate solution.

Now we consider cases where tax distortion is non-negligible and a National Treatment clause limits a host's ability to obtain sufficient compensation for bearing all the post-investment risk. In this case, $N < \infty$ is efficient, while a narrow police powers carve-out might limit investment and also be inequitable. In order to make the point simply, we continue to assume $T < \tilde{T}$.¹² We begin by showing that the contract curve is vertical when conditions C1-C4 all hold.

Proposition 3 *When conditions C1, C2, C3, and C4 hold, the host's and source's indifference curves are tangent at a unique tax rate T^* . The contract curve consists of a vertical segment*

$$\left\{ T, N : T = T^*, N \in \left(\frac{H \ln \frac{K}{C}}{H + R - K - T^*}, \frac{K - \alpha C}{2} \left[\frac{D''(T^*) - 1/H}{\frac{T^* + K - R}{H} - D'(T^*)} \right] \right) \right\}.$$

The proof contains the implicit equation that determines the value of T^* . Proposition 3 provides sufficient conditions for tangency of the source's and host's indifference curves at interior values of N . Importantly, pairs (T, N) along the resulting contract curve are locally efficient. In short, when there is a non-trivial bargaining inefficiency, a broad police powers carve-out can be locally efficient. Convexity of the preference sets is guaranteed only for a subset of the relevant choice set, so we cannot establish global efficiency.

¹²Even though T is a choice variable, restricting its range is valid: certain parameter and functional form restrictions imply $T < \tilde{T}$ in any sensible agreement. For example, if we restrict parameter values such that $D'(\tilde{T}) < 1$ and $D'' \geq 0$, then for any $T \geq \tilde{T}$, $\frac{\partial E(W_H)}{\partial T} = \beta_2 - D'(T) < 0$. Since no sensible contract will specify a transfer payment that hurts both the source and the host, contracts with $T \geq \tilde{T}$ will not be viable.

The contract curve is a vertical segment at a unique transfer payment T^* . The uniqueness of T^* means it is not possible to raise the source's welfare by raising N and still compensate the host for bearing more of the ex post risk associated with investment. Instead, a larger value of N (together with an efficient tax T^*) represents a transfer of welfare from the host to the investor. In this situation, the claim that a strict interpretation of Chapter 11's expropriation clause benefits investors at the expense of host nations is correct. Moreover, as we show below, under some circumstances IIAs with high N s can also inhibit investment.

4.3 Investment

In the second stage, after the investment has been sunk, the project might produce negative expected benefits to either the host or the investor. In the first stage, the investor has the option of walking away from the project, so the investor's ex ante expectation of its benefit from the project is non-negative in any first stage equilibrium. The host would like to reject projects that generate negative expected benefits, but its ability to do so depends on the terms of the IIA.

To see how a narrow police powers carve-out may reduce investment, consider the case where the IIA and judicial interpretation lead to a narrow police powers carve-out (large N), a project is risky (large H), and the marginal deadweight loss, D' is large. In this situation the host requires a large compensation ($T - D(T)$) in order to benefit from the project, i.e. in order that $E(W_H) \geq 0$. If the host is able to reject projects, the narrow police powers carve-out may result in the rejection of (potentially) mutually beneficial projects. If the host cannot reject projects, the narrow police powers carve-out can lead to expected losses for the host.

For example, suppose that the source has all of the bargaining power, and the host must receive expected welfare no less than that associated with the indifference curve labeled W in Figure 2 (i.e. the host has some ability to reject projects). If the source could choose both T and N , the outcome would be T^* and N^* (where the source's indifference curve S^* is tangent to the host's participation constraint, W). However, if N is set exogenously at $\hat{N} > N^*$, then the source needs to offer the host \hat{T} , the level that satisfies the host's participation constraint. At \hat{N}, \hat{T} , the source obtains a level of welfare given by the indifference curve \hat{S} . If this welfare level is below the source's reservation level, the host and source will be unable to arrive at a mutually beneficial agreement regarding the terms of investment.

Another possibility (not drawn) is that N is predetermined at a level N^1 so high that there is no value of T that can guarantee the host non-negative expected welfare, so the host rejects the

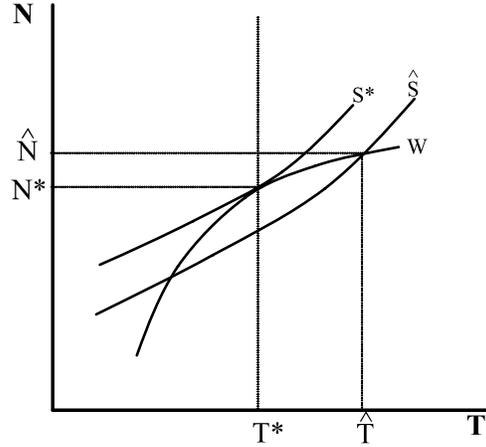


Figure 2: Host's participation constraint, (W) and investor's indifference curves (S).

project. This is possible when condition C2 is violated, i.e. when the marginal tax distortion is so large that, at high T , it swamps any direct expected benefits from the project.

Most IIAs provide only weak rights to invest and hosts are able to reject unattractive projects. For these agreements, the greater regulatory risk caused by a narrow police powers carve-out might lead hosts to reject many projects, reducing the equilibrium level of FDI. NAFTA is more complicated. Rights of establishment and acquisition are included in the MFN and National Treatment clauses, making it difficult for the host to reject risky investments. NAFTA hosts may have to accept projects that they expect will not be beneficial for them. However, NAFTA signatories had the option to exempt entire sectors from the MFN and National Treatment clauses. To the extent that the signatories were able to anticipate (at the time of signing) which sectors would be risky, they do not need to accept projects with negative expected benefits. However, exempting these sectors means that they enjoy none of the positive effects of the agreement (increased security and transparency).

4.3.1 A numerical example

Here we provide a numerical example of the impact that a strict expropriation clause may have in the context of an IIA with weak rights to invest: the project is rejected by the host when N is high, but accepted by both host and source when N is lowered sufficiently.

Assume that $D(T)$ is quadratic: $D(T) = \max\{0, aT^2/2 - bT\}$. We use the following parameter values in our example: $\pi = 2$, $H = 1.5$, $C = 0.1$, $R = 0.8$, $K = 1$, $a = 1$, $b = .1$.

Then $D(T)$, $D'(T)$ and $D''(T)$ are positive whenever $T > 0.2$. By equation (10) of Appendix A, $T^* = 0.475$, which is less than \tilde{T} provided $N > 4.189$. Moreover, $\frac{K+T^*-R}{H} - D'(T^*) = 0.075$ while $\bar{T} \equiv \frac{R+\pi}{2} - K = .4$, rendering the host's welfare increasing and the source's welfare decreasing in T at $T = T^*$. The host's preferences are convex at T^* for some N (using equation (9) in the appendix); in particular, the host's preference set is convex provided $N < 19.922$. In sum, under the parameter values above, the host's and source's indifference curves are tangent at $T^* = 0.475$ provided $N \in (4.189, 19.922)$. We note that, because $T^* < R$, the holdup problem would be present in this example if there were no compensation rule in place.

Finally, substituting values into (4) gives $E(W_H|T = T^*) = -0.208 + \frac{1.130}{N}$ and setting this expression to 0 implies $N^* = 5.906$. By (5), at (T^*, N^*) the source's expected payoffs (net of up front costs K) equals 0.084. Thus, both the host and the source are willing to undertake the project when $(T, N) = (T^*, N^*)$. Suppose that, instead, the IIA pegs N at $N^1 = 7$. At N^1 , there is *no* transfer that can ensure the host non-negative expected welfare, and the host will reject the project outright. Thus, in this example, a sufficiently high value of N *hinders* potentially mutually beneficial investment.

5 Conclusion

A narrow police powers carve-out to the expropriation clause in NAFTA's Chapter 11 can promote investment by helping to overcome the hold-up problem and the tendency of governments not to internalize regulatory costs. Some argue that governments should be exempt from paying compensation for regulations that serve a legitimate public purpose and that are non-discriminatory. Most NAFTA tribunals have been sympathetic to such concerns and have applied a police powers carve-out to expropriation.

We used a simple model to analyze the efficiency and equity implications of a police powers carve-out, given the presence of non-discrimination rules. A police powers carve-out creates two inefficiencies during the post-investment (regulation) stage. First, when parties can dispute whether a regulation is legitimate, a carve-out may encourage costly litigation. Second, a positive probability of a tribunal exempting the regulator from paying compensation encourages the host to regulate more often than is warranted from a joint welfare perspective. Both of these inefficiencies stem from moral hazard. If these are the only inefficiencies in the investment game, then there is no role for a police powers carve-out. It is Pareto efficient to adopt a narrow

police powers carve-out, and use taxes that are set prior to investment to compensate the host country for the increased regulatory risk.

However most modern investment treaties also potentially generate inefficiency at the stage when the terms of the investment are determined. For example, the National Treatment principle requires the host to treat foreign investors no less favorably than domestic investors. A host government cannot demand side payments from foreign investors, nor can it charge them higher tax rates. This aspect of National Treatment can create inefficiencies when the investment project is environmentally risky. In order to be willing to accept a risky project while surrendering the right to regulate the project without paying compensation, the host may need a significant payment from the investor. The host may be unable to obtain this payment and still comply with the National Treatment clause, unless it imposes an inefficiently high tax on domestic investment.

Distortions associated with non-discrimination clauses are central in our analysis. We identified conditions under which this first stage inefficiency dominates the moral hazard problems surrounding regulation. When the tax distortion is sufficiently important, a range of police powers carve-outs are all Pareto-efficient. A narrow carve-out might simply transfer rent from the host government to foreign investors.

We provided an example in which narrowing the police powers carve-out can reduce investment when signatories do not have a strong right of establishment. Even in NAFTA (which does give a such a right) parties were able to select which sectors to exempt from aspects of the treaty. When hosts can reject projects, investments must satisfy a participation constraint for both the host and the investor. If the police powers carve-out to the expropriation clause is narrow, the host requires a large payment to accept risky projects. In some cases, there may be no outcome that satisfies both parties' participation constraints. In such a case, a broader police powers carve-out would increase the range of mutually beneficial investment opportunities. A narrow police powers carve-out might inhibit rather than encourage FDI.

Appendix A: Proofs

Proof of Proposition 1 The relation $i \prec j$ means that action j gives the host a higher payoff than action i . By Table 1, $b \prec c$ if and only if $\beta > \beta_1$, where β_1 solves the relation $P(\beta_1, N)K = C$. Also by Table 1, $a \prec c$ if and only if $\beta > \beta_2$ where β_2 solves

$R - (1 - P(\beta_2, N))K - C = T - \beta_2 H$. Similarly, $a < b$ if and only if $\beta > \beta_3$, where β_3 solves $T - \beta_3 H = R - K$. Finally, $\beta_1 \begin{cases} > \\ \leq \end{cases} \beta_2 \begin{cases} > \\ \leq \end{cases} \beta_3$ if and only if $1 - \frac{\ln \frac{K}{C}}{N} \begin{cases} > \\ \leq \end{cases} \frac{T+K-R}{H}$, which holds if and only if $T \begin{cases} < \\ \geq \end{cases} \tilde{T}$. Provided the probability distribution for β does not have a mass point at one of the critical values, we do not need to specify which action the host will take when indifferent between two options.

Proof of Lemma 1 When condition C1 holds, $E(W_H)$ and $E(W_S)$ are as given by equations (4) and (5) and the corresponding partial derivatives are as given in Table 2. Using values found in Table 2, we see that $dE(W_H)/dT > 0$ if and only if C2 holds. The maintained assumption $K > C$ implies that $\alpha C < K$ which in turn implies that $dE(W_H)/dN$ is negative. Thus, conditional on C1, $\frac{dN}{dT} \Big|_{dE(W_H)=0}$ is positive if and only if C2 holds.

Condition C3 is necessary and sufficient for $dE(W_S)/dT > 0$. Our maintained assumption $K > C$ ensures that $dE(W_S)/dN > 0$. Thus, given C1, $\frac{dN}{dT} \Big|_{dE(W_S)=0} > 0$ if and only if C3 holds.

Proof of Lemma 2 Condition C1 implies $E(W_H)$ and $E(W_S)$ are given by equations (4) and (5) respectively and the corresponding partial derivatives are as given in Table 2. From Table 2, the cross partial derivatives of both welfare functions are 0, so the functions are additively separable in T, N . For both agents, we can write expected utility as $G(T) + F(N)$. Of course, the functions are different for the two agents, but to conserve notation we avoid subscripts. For the host, $G' > 0, F' < 0$ and these inequalities are reversed for the source. It is easy to show that the formula for the second derivative of an indifference curve is

$$\frac{d^2 N}{dT^2} = \left(- (F')^2 G'' - (G')^2 F'' \right) \frac{1}{(F')^3} \quad (8)$$

Using Table 2, for the source, we have

$G' = -\frac{(K-R+T) - (\pi-K-T)}{H}$	$G'' = \frac{d\left(-\frac{(K-R+T) - (\pi-K-T)}{H}\right)}{dT} = -\frac{2}{H}$
$F' = \frac{K-C}{N^2}$	$F'' = 2\frac{-K+C}{N^3}$

Table A1. First and Second derivatives of source welfare

Given our maintained assumption that $K > C$, we see from Table A1 that for the source $F' > 0$. Thus the cubic that multiplies the large parentheses in equation (8) is positive, so the sign of $\frac{d^2 N}{dT^2}$ is the same as the sign of

$$- \left((F')^2 G'' + (G')^2 F'' \right)$$

We can also see from Table A1 that F'' and G'' are always negative. Since $(G')^2$ is always positive, condition C1 is sufficient to ensure that $\frac{d^2 N}{dT^2} > 0$ for the source. As shown in the proof of Lemma 1, given C1 holds, condition C3 is necessary and sufficient to ensure that the source's welfare is increasing as N becomes larger and T becomes smaller. Thus the source's preferred sets are convex when C1 and C3 hold.

The first and second derivatives of the host's welfare are

$G' = \frac{-R+K+T}{H} - D'$	$G'' = \frac{1}{H} - D''$
$F' = \frac{\alpha C - K}{N^2}$	$F'' = -2 \frac{\alpha C - K}{N^3}$

Table A2. First and Second derivatives of host welfare

Since (for the host) $F' < 0$, the cubic that multiplies the large parentheses in equation (8) is negative. Thus the sign of $\frac{d^2 N}{dT^2}$ is the same as the sign of

$$(F')^2 G'' + (G')^2 F'' \tag{9}$$

All the terms in equation (9) are always positive except for G'' . Thus $\frac{d^2 N}{dT^2}$ is negative if and only if D'' is sufficiently large, i.e. if condition C4 holds.

Convexity of the hosts preferences requires that either $\frac{dN}{dT} \Big|_{dE(W_H)=0} > 0$ and $\frac{d^2 N}{dT^2} < 0$ or $\frac{dN}{dT} \Big|_{dE(W_H)=0} < 0$ and $\frac{d^2 N}{dT^2} > 0$. Thus with reference to Lemma 1 we see that convexity requires that either both C2 and C4 hold, or that neither hold.

Proof of Proposition 2 Condition C1 implies $E(W_H)$ and $E(W_S)$ are as given by (4) and (5). By Lemma 1, the host's and the source's indifference curves slope upward (N as a function of T) when conditions C1-C3 hold. However, when $D' \approx D'' \approx 0$, the host's indifference curve is convex (N as a function of T) and, when $\pi - T - K \geq 0$, is *steeper* than the source's. In this situation, any N^o, T^o satisfying $T \in (\bar{T}, \tilde{T})$ can be improved upon by moving upward along one agent's indifference curve (holding that agent's welfare constant) and increasing both T and N .

Proof of Proposition 3 Condition C1 implies $E(W_H)$ and $E(W_S)$ are as given by (4) and (5). When conditions C1-C4 hold, the indifference curves slope upward and each agents' preferred set is convex. Under these conditions, the contract curve is the set of points in the T, N plane on which the indifference curves are tangent. The slopes of both indifference curves are proportional to N^2 , implying that the contract curve is independent of N ; that is, the contract curve is vertical at the unique tax, denoted T^* , at which the slopes in equations (6) and (7) are equal to each other. The equation for T^* is:

$$\frac{K - R + T^* - D'(T^*)H}{K - \alpha C} = \frac{K - R + T^* - (\pi - T^* - K)}{K - C}. \quad (10)$$

As noted above, points along the vertical line at T^* form the contract curve only if preferences are convex. From Lemma 2, conditions C1, C2 and C4 jointly guarantee a convex preferred set for the host. From Table 3 we see that condition C4 requires $N < \frac{K - \alpha C}{2} \left[\frac{D''(T^*) - 1/H}{\left[\frac{T^* + K - R}{H} - D'(T^*) \right]^2} \right]$. Finally, recalling equation (2), $T^* < \tilde{T}$ if and only if $N > \frac{H \ln \frac{K}{C}}{H + R - K - T^*}$.

Appendix B: Indifference curves when $T \geq \tilde{T}$

When $T \geq \tilde{T}$ then $\beta_3 \geq \beta_2 \geq \beta_1$ and the host only undertakes two possible actions in the second stage: (a) do not regulate, or (c) regulate and litigate. The host does not regulate and compensate because the tax payment received in the absence of regulation is too attractive to pass up unless there's a high likelihood of winning in court. When $T \geq \tilde{T}$, ex ante expected welfare for the host and source are given by

$$E(W_H) = \int_0^{\beta_2} [T - \beta H] d\beta + \int_{\beta_2}^1 [R - [1 - P(\beta; N)]K - C] d\beta$$

and

$$E(W_S) = \int_0^{\beta_2} [\pi - T] d\beta + \int_{\beta_2}^1 K [1 - P(\beta; N)] d\beta.$$

Differentiating and rearranging gives

$$\left. \frac{dN}{dT} \right|_{dE(W_H)=0} = \frac{\beta_2 - D'(T)}{aK} \quad \text{and} \quad \left. \frac{dN}{dT} \right|_{dE(W_S)=0} = \frac{\beta_2 - \gamma \frac{\partial \beta_2}{\partial T}}{\eta K + \gamma \frac{\partial \beta_2}{\partial N}}$$

where $\eta \equiv \int_{\beta_2}^1 P(\beta; N) d\beta > 0$ and $\gamma \equiv \pi - T - [1 - P(\beta_2; N)]K$. Since $b > 0$ is a necessary condition for the source to invest, and $\partial \beta_2 / \partial T > 0$, $\partial \beta_2 / \partial N > 0$, then whenever $D' \approx 0$, the

host's indifference curve is steeper than the source's. Thus, just as when $T < \tilde{T}$ as described in Proposition 2, if $D' \approx 0$, any contract (N^o, T^o) with $N^o < \infty$ and $T^o \geq \tilde{T}$ can be improved upon by raising N and compensating the host via correspondingly higher T .

References

- ALVAREZ, G. A., AND W. W. PARK (2003): “The New Face of Investment Arbitration: NAFTA Chapter 11,” *Yale Journal of International Law*, pp. 365–403.
- BAGWELL, K., AND R. STAIGER (2001a): “Domestic Policies, National Sovereignty and International Economic Institutions,” *Quarterly Journal of Economics*, pp. 519–62.
- (2001b): “The WTO as a mechanism for securing market access property rights: implications for global labor and environmental issues,” *Journal of Economic Perspectives*, 15(3), 69–88.
- BEEN, V., AND J. C. BEAUVAIS (2003): “The Global Fifth Amendment? NAFTA’s Investment Protections and the Misguided Quest for an International ‘Regulatory Takings’ Doctrine,” *New York University Law Review*.
- BHALA, R. (1999): “The Myth about Stare Decisis and International Trade Law,” *American University International Law Review*, 14, 845–956.
- BLUME, L., D. L. RUBINFELD, AND P. SHAPIRO (1984): “The taking of land: When should compensation be paid?,” *Quarterly Journal of Economics*, pp. 71–92.
- BRITISH COLUMBIA SUPREME COURT (2002): “Reasons for Judgment of May 2, 2001, The United Mexican States v. Metalclad Corporation, 2001 BCSC 664,” *I.L.R.*, 647, 119.
- CANADA DEPARTMENT OF FOREIGN AFFAIRS (2004): “Statement of Claim, Government of Canada Statement of Defense,” http://www.dfait-maeci.gc.ca/tna-nac/disp/ethyl_archive-en.asp.
- CHUA, A. (1998): “The Precedential Effect of WTO Panel and AB Reports,” *Leiden Journal of International Law*, 11, 45–61.
- DE MOOIJ, R. A., AND S. EDERVEEN (2003): “Taxation and foreign direct investment: a synthesis of empirical research,” *International Tax and Public Finance*, (3), 673 – 693.
- EDERINGTON, J. (2001): “International coordination of trade and domestic policies,” *American Economic Review*, 91(5).

- ENVIRONMENT CANADA (1998): “News Release,” http://www.ec.gc.ca/press/mmt98_n_e.htm.
- EPSTEIN, R. A. (1985): *Takings: Private Property and the Power of Eminent Domain*. Harvard University Press, Cambridge, MA.
- GAINES, S. (2002): “The Masked Ball of NAFTA, Chapter 11,” in *Linking Trade, Environment and Social Cohesion*, ed. by J. Kirton, and V. Maclaren. Ashgate.
- GLAMIS GOLD LTD. (2003): “Notice of Arbitration,” <http://www.state.gov/documents/organization/27320.pdf>.
- GLOBAL LEAD NETWORK (2005): “MMT Factsheet,” <http://www.globalleadnet.org/advocacy/initiatives/mmtfactsheet.cfm>.
- GREIDER, W. (2001): “The Right and US Trade Law: Invalidating the 20th Century,” *The Nation*.
- GROSSMAN, S., AND O. HART (1986): “The costs and benefits of ownership: a theory of vertical and lateral organization,” *Journal of Political Economy*, pp. 691 – 719.
- HALLWARD-DRIEMEIER, M. (2003): “Do Bilateral Investment Treaties Attract FDI? Only a bit and they could bite,” World Bank Policy Research Paper WPS 3121.
- HUANG, Y. (2005): “Are Foreign Firms Privileged by their Host Governments? Evidence from the 2000 World Business Environment Survey,” MIT Sloan Working Paper No. 4538-05.
- INTERNATIONAL CENTRE FOR SETTLEMENT OF INVESTMENT DISPUTES (2000): “Metal-clad Corp. v. United Mexican States (Award), ICSID Case No. ARB(AF)/97/1 para. 103 (Aug. 30, 2000),” .
- LEVINSON, A. (1997): “A note on environmental federalism: interpreting some contradictory results,” *Journal of Environmental Economics and Management*, 33, 359–366.
- MANN, H. (2001): “Private Rights, Public Problems: a Guide to NAFTA’s controversial Chapter on Investors Rights,” International Institute on Sustainable Development, Winnipeg.
- MARKUSEN, J. (2001): “Commitment to rules on investment: The developing countries’ stake,” *Review of International Economics*, 9, 287–302.

- NEUMAYER, E., AND L. SPESS (2005): “Do Bilateral Investment Treaties Increase Foreign Direct Investment to Developing Countries?,” *World Development*, 33(10), 1567–1585.
- NOLDEKE, G., AND K. SCHMIDT (1995): “Sequential investments and options to own,” *Rand Journal of Economics*, pp. 633 – 53.
- PANAGARIYA, A. (1999): “The Millenium Round and Developing Countries: Negotiating Strategies and Areas of Benefits,” in *Conference on Developing Countries and the New Multilateral Round of Trade Negotiations*, Harvard University.
- PRICE, D. M. (2000): “Some Observations on Chapter Eleven of NAFTA,” *Hastings International and Comparative Law Review*, pp. 421–425.
- Public Citizen (2001): “NAFTA Chapter 11 Investor to State cases: Bankrupting Democracy,” <http://www.citizen.org/publications/release.cfm?ID=7076>.
- SCHMITZ, P. (2001): “The hold-up problem and incomplete contracts: a survey of recent topics in contract theory,” *Bulletin of Economic Research*, pp. 1 – 17.
- SOLOWAY, J. (2002): “Environmental expropriation under NAFTA Chapter 11: the Phantom Menace,” in *Linking Trade, Environment and Social Cohesion*, ed. by J. Kirton, and V. Macclaren. Ashgate.
- TOBIN, J., AND S. ROSE-ACKERMAN (2004): “Foreign Direct Investment and the Business Environment in Developing Countries: The Impact of Bilateral Investment Treaties,” Yale Law School Center for Law, Economics and Public Policy Research Paper No. 293.
- TSCHEN, A. (1999): “Chapter 11: The Efforts to Define Expropriation,” *Currents: International Trade Law Journal*, pp. 50–68.
- TURK, J. (2005): “Compensation for “Measures Tantamount to Expropriation” under NAFTA: What it Means and Why it Matters,” *International Law & Management Review*, pp. 41–62.
- UNCITRAL TRIBUNAL METHANEX CORP. V. UNITED STATES OF AMERICA (2005): “Final Award of the Tribunal on Jurisdiction and Merits,” <http://www.state.gov/documents/organization/51052.pdf>.

UNCTAD (2005): “Latest Developments in Investor-State Dispute Settlement,” IIA Monitor Number 4.

UNITED STATES OF AMERICA (2005): “Statement of Defense,” <http://www.state.gov/documents/organization/45118.pdf>.

U.S DEPARTMENT OF STATE (2005a): “Summary of Glamis Gold Ltd. vs United States of America,” <http://www.state.gov/s/l/c10986.htm>.

——— (2005b): “Summary of Metalclad Corp. v. United Mexican States,” <http://www.state.gov/s/l/c3752.htm>.

U.S. DEPARTMENT OF STATE (2005): “Summary Methanex Corp. v. United States of America,” <http://www.state.gov/s/l/c5818.htm>.

WARNER, M. (2005): “Analysis of Ethyl Inc. v. Canada,” <http://government.cce.cornell.edu/doc/pdf/EthylAp.pdf>.

WILSON, J. (1997): “Capital mobility and environmental standards: Is there a theoretical basis for a race to the bottom?,” in *Fair Trade and Harmonization: Prerequisites for Free Trade?*, ed. by J. Bhagwati, and R. Hudec. MIT Press.